ABSTRACT

SATIROGLU, SINEM. Genetic Parameters Estimates from *Pinus taeda L.* Fourth Cycle Coastal Breeding Population (Under the direction of Dr. Fikret Isik).

In this study, genetic parameters, and breeding values of 294 parents were estimated in *Pinus taeda L*. (loblolly pine) for growth traits (height and stem diameter), stem forking defects and fusiform rust disease, incited by the fungus *Cronartium quercuum* f. sp. *fusiforme*. Data from the 4th-Cycle Coastal Breeding population managed by the Cooperative Tree Improvement Program at North Carolina State University were used for all analyses. Linear mixed models with various variance-covariance structures were implemented to estimate variance components in 25 multi-environmental trials.

Additive genetic variance explained 56 and 67% of genetic variation for stem diameter and height, respectively. Additive genetic variance was 1.3 and 2.1 times greater than the dominance genetic variance for stem diameter and height, respectively. The individual-tree narrow-sense and broad-sense heritabilities for height were 0.15 and 0.22, respectively. For stem diameter, the two heritabilities were 0.09 and 0.17. For height, the additive GxE correlation averaged 0.69. The correlation was 0.56 for stem diameter, indicating considerable GxE interaction for both traits.

For fusiform rust disease incidence, additive genetic variance explained 78% of genetic variance, which was 3.7 times higher than the dominance variance. For stem forking, additive and non-additive genetic variances were low. The narrow-sense and broad-sense heritabilities for fusiform rust disease incidence were 0.25 and 0.31, respectively. In stem forking, the heritability estimates were essentially zero, indicating that the stem forking is mostly controlled by environmental factors.

Genetic Parameter Estimates from *Pinus taeda* L. Fourth Cycle Coastal Breeding Population

by
Sinem Satiroglu

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Forestry and Environmental Resources

Raleigh, North Carolina
2021

APPROVED BY:

_____	_____
Dr. Fikret Isik Committee Chair	Dr. Justin G. A. Whitehill
_____	
Dr. Steven E. McKeand	

DEDICATION

For my family, my mother and father Süreyya and Osman and my brother Servet, for always supporting and believing in me. Thank you for raising me as confident to have big dreams and me in achieving dreams. For all of my relatives, who showed their love and support when I needed it.

For my beloved one, Yusuf Ali Ergin who completes me. Thank you for being with me all these years. Thank you for your patience, support, and love.

For my best friend, Hatice Duygu Bankoğlu, thank you for all the support and motivating talk when I struggle. As two friends who have similar dreams, you are the one who understands me well.

Thank you for all of your support to Selen, Uğur, and all of my friends who I met through my journey and colleagues who helped me whenever and wherever. You will always be in my heart.

BIOGRAPHY

Sinem Satiroglu was born on 2 September 1990 in Trabzon, Turkey, and grew up in the same city. Sinem obtained her bachelor's degree in Forestry Engineering from the Faculty of Forestry, Black Sea Technical University, Turkey in 2012. Sinem started her M.S. degree at the same university between 2012 to 2015 studying "Estimation of Aboveground Biomass in Oak Coppice in Arapgir Region". In 2016, she was awarded a full M.S. scholarship by the Ministry of Natural Education of Republic of Turkey to study in the United States. She started her M.S. in the Department of Forestry and Environmental Resources at North Carolina State University, Raleigh, NC.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Fikret Isik for all his invaluable guidance and expertise. Thank you for always being kind and patient and being there whenever I had questions or needed just for a talk. I am very thankful to my committee members Dr. Steven McKeand and Dr. Justin G. A. Whitehill for their patience and guidance. I would like to thank Dr. Nasir Shalizi for his effort, patience, and help. He spent a significant amount of time helping me finish this study. I would like to thank my previous advisor Dr. Kitt Payn for his guidance. I also would like to thank the North Carolina State University Cooperative Tree Improvement Program for supporting my study. Trevor Walker, Austin Heine, Dr. Eddie Lauer, and all The Cooperative Program staff and graduate students were always supportive. I would like to thank the Republic of Turkey for funding me with a scholarship to achieve this degree in the US.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Estimation of Genetic Parameters from Loblolly pine (*Pinus taeda* L.) Progeny Tests

1.1. Introduction

The demand for wood and wood products is increasing in the world (Bowyer, 2000; Turner et al., 2006). Managed forest plantations can play an important role to meet the demand in a sustainable way, but to increase wood productivity, forest plantations need to be supported by tree breeding and silvicultural systems. The model in the southern USA is a great example for sustainable wood production from intensively managed tree farming while preserving natural forests (McKeand et al., 2021). The southeastern United States pine breeding programs have shown that investment in tree breeding is worthwhile due to the gains made over non-improved seedlings (McKeand et al., 2006). For example, the estimated genetic gain in growth for open-pollinated families is almost 50% higher compared with non-improved seedlings (McKeand, 2019). These gains are made through progeny testing, selection, and breeding.

The time span needed for a breeding cycle is long for forest trees because the testing and selection phases typically continue over a decade. The breeding process is logistically complex and costly. Sexual maturation of trees to flower and phenotypic expression of trees in progeny tests takes a long time (White et al., 2007). When these difficulties are taken into account, well detailed plans and mating designs are needed to manage the complexity, control the inbreeding level, and increase genetic gain in multiple traits (Isik & Mckeand, 2019).

Progeny testing is the most important and logistically challenging component of forest tree breeding. Progeny testing is considered the backbone of tree breeding and includes the: creation of progenies; establishment of experiments with a family structure; statistical evaluation; and genetic interpretation of results and breeding decisions. It requires resources to establish, manage and assess many field trials after a certain number of years. The generation of progenies, the establishment of experiments with a family structure, statistical assessment, genetic interpretation of data, and breeding decisions are all part of the progeny testing process (Lindgren, 1991).

Progeny tests provide data to predict the genetic merit of parents and their progeny (Mili et al., 2010). The breeding value (BV) of a tree is typically determined by the mean phenotype of its progeny from random mating among parents in the population (Falconer & Mackay, 1996). The breeding values (genetic merit) are used to rank individuals, make selections, and establish the next cycle of the breeding population (Ogut et al., 2014). Progeny testing is also known as genetic testing of an individual breeding value's offspring (Slamet, 2020). Breeders can determine breeding values of families or parents with high precision using phenotypic data collected from progeny testing.

Progeny test results are informative to understand the inheritance of complex traits, such as additive and non-additive genetic sources of variability and heritability (Isik & Mckeand, 2019). Most traits of interest to tree breeders are controlled by many genes each contributing minor effects to the overall phenotype (Isik et al., 2017). Estimation of genetic parameters is a vitally important step in understanding the degree of genetic control and to predict genetic gain in a breeding population (White et al., 2007; Piepho & Williams, 2010). These statistics inform the breeder about

the degree to which a trait might be inherited (heritability), and the degree to which any two traits are genetically correlated (White et al., 2007). Producing these statistics to characterize a breeding population is necessary for making decisions for a selection and testing regime and to make gain in successive iterations of the breeding and testing process (Zobel & Talbert, 1984; Falconer & Mackay, 1996; Sims, 2017). When the heritability estimates are low, a relatively low proportion of phenotypic variance can be attributable to genetic causes. For low heritability traits, the phenotype of an individual is often a poor predictor of its genetic merit because of large environmental effects. If this is the case, parental selection based on progeny performance is more reliable (Slamet, 2020).

In forest tree progeny testing, selections are usually tested in a range of environments to draw general conclusions for a deployment zone (Lauer et al., 2021). Even if we use the same genetic material for a progeny test, phenotypic expressions may be different, because of environmental factors. Genotype by environment interaction (GxE) modeling can be used to understand the norm of reaction and relative ranking of families at different environments (Xiong et al., 2010). Therefore, breeding efforts should be focused on certain geographic areas where plantations have already been established. Deciding on the area where genetic material will be used is one of the goals of progeny testing. Replications of the progeny test should be dispersed over geographic regions big enough to enable such judgments for this purpose.

Multi-environmental trials (MET) allow researchers to evaluate genotypes in different environmental conditions, to compare their responses in those environments and their stability and adaptability. One of the most significant functions of MET is to enable investigation of genotype

by environment (GxE) interactions. MET studies help breeders select the best genotypes across environments (Cullis et al., 1998; Smith et al., 2005; Ogut, 2012).

The analysis of progeny tests can be a challenge when the data are not balanced (not all the genotypes are tested in all environments) (Isik et al., 2005). The MET data in forest tree progeny tests usually have heterogeneous variance and covariance structures (Smith et al., 2001; Crossa, 2006; Ogut, 2012). Historically, weak genetic connections between tests is a common problem in forest tree breeding due to unbalanced data structures (Piepho et al., 2008; So, 2009; Ogut, 2012). Analyzing MET data can be statistically complex due to the incomplete and unbalanced data (Spilke, 2005; Kelly, 2007; Ogut, 2012).

In recent years, incomplete block designs have been popular in forest tree progeny tests to accommodate large number of entries. These designs, such as alpha-cyclic incomplete row-column, increase testing efficiency (John & Williams, 1995). These designs allow for controlling the environmental gradients at a location in two directions (rows and columns) to better control the environmental noise (White, 1996; Isik & Mckeand, 2019). The partitioning of degrees of freedom into extra blocking strata, as well as the replication effect, contribute to the efficiency of resolvable row-column designs. For making more accurate selection decisions, connecting progeny tests established in different years or in different series is crucially important. This way, thousands of genetic entries can be compared for selection. It is suggested computer generated incomplete block designs is an efficient way to accommodate a large number of treatments while controlling the environmental noise in genetic tests (John & Williams, 1995). Spatial row-column

design is a modified version of incomplete block designs. It allows fitting spatial residual structures in progeny test data analyses.

Incomplete block designs (IBD) are a type of block design that can be employed when block sizes are big (or any time micro-environments within a block become heterogeneous in field, greenhouse, nursery, or growth room situations). Each full block (also known as a resolvable replication) is subdivided into smaller parts known as incomplete blocks or simply blocks with IBDs (Cochran & Cox, 1957; Williams & Matheson, 1994; John & Williams, 1995). Each incomplete block comprises a portion of the genetic entries, which are planted in a more consistent micro-environment than would be conceivable if all entries were planted in a single complete block. The incomplete blocks include a new blocking factor that is used to compensate for micro-site disparities in entry ways. IBDs improve the precision of genetic entry comparisons by removing variation between incomplete blocks and block-to-block variation, as well as lowering experimental error. In several forestry scenarios, IBDs have proven to be effective for genetic testing (Williams & Matheson, 1994). The alpha-lattice design is a sort of IBD that is very effective and adaptable for genetic tests with many entries (Patterson & Williams, 1976; Patterson et al., 1978). Such designs can be easily implemented using computer algorithms such as ALPHA+, ALPHAGEN (Williams & Talbot, 1993) and CycDesigN (Whitaker et al., 2001).

1.2 Progeny Testing in 4th-Cycle Loblolly Pine

The Cooperative Tree Improvement Program at NC State University aims to improve *P. taeda* for growth, stem quality, and disease resistance in the southern USA. For this purpose, the program has created partnerships among the university, forest companies and state agencies for over 65 years. The program's main goal is to increase value to landowners by creating highly productive forests (McKeand et al., 2021). The breeding program rigorously tests the trees and selects elite individuals that offer the greatest potential value, all the while maintaining genetic diversity for future breeding generations.

For the 4th-cycle progeny testing, the Cooperative used the rolling front progeny testing technique in the fourth cycle to manage large numbers of progeny testing activities spanning across years. When enough seed of new genetic entries (> 70) were available for the year, progeny tests were established. Planting 60 offspring for a cross was regarded as sufficient to predict breeding values of parents with a specified accuracy of 0.80 based on sensitivity analysis of internal data sets (unpublished results). Families were typically tested over two years, with 30 seedlings tested in year one and the remaining 30 seedlings tested in year two. The overlapping of entries among test series was accomplished by dividing the progeny of a family across years. To improve the accuracy of contrasts across families throughout test series, some crosses examined in a year were included in multiple year test series (Isik & Mckeand, 2019).

1.3 Linear mixed models for progeny tests data analysis

Estimating genetic parameters such as additive, non-additive, phenotypic variance, heritability, and breeding values is important stage of breeding process. Through these estimations, selection can proceed, and the genetic gain can be estimated (Zobel & Talbert, 1984; Falconer & Mackay, 1996).

Mixed models are the standard statistical approaches used to analyze progeny tests data. They are flexible to model complex variance-covariance structures and can handle unbalanced or missing data (Smith et al., 2005). A general linear mixed model can be expressed in matrix form as

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e} \tag{1.1}$$

Where **y** is a vector of observations, **β** is a vector of fixed effects with design matrix **X**, **u** is a vector of random effects with its design matrix **Z**, **e** is a vector of residual errors. The mean, site, replicate, and covariates can be fixed effects, whereas random effects may include incomplete block and plot effects, as well as tree of family effect (breeding values) (Dutkowski, 2005). The regular assumptions of mixed models can be summarized as follows.

$$\boldsymbol{V} = \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \tag{1.2}$$

Where **V** is the variance-covariance of observations (vector **y**), **G** is the variance-covariance of random effects **u**, and **R** is the variance-covariance of residuals **e**. In the simplest cases, we assume

the **G** and **R** matrices are $\mathbf{G} = \mathbf{I}\sigma_u^2$ and $\mathbf{R} = \mathbf{I}\sigma_e^2$, where **I** is the identity matrix, $\boldsymbol{\sigma_u^2}$ is the variance of random effects and $\boldsymbol{\sigma_e^2}$ is the variance components of residuals. The mixed model equations (MME) for general linear mixed can be presented as follows (Henderson, 1984);

$$\begin{bmatrix} \mathbf{X`\,R^{-1}\,X} & \mathbf{X`\,R^{-1}\,Z} \\ \mathbf{Z`\,R^{-1}\,X} & \mathbf{Z`\,R^{-1}\,Z + G^{-1}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{b}} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X`\,R^{-1}\,y} \\ \mathbf{Z`\,R^{-1}}\boldsymbol{y} \end{bmatrix} \tag{1.3}$$

G is the direct sum of the variance/covariance matrices of each of the random effects, where **R** is the variance/covariance matrix of the residuals. The random effect terms are considered to be independent (Mrode, 2014).

The variance/covariance matrices in **G** are usually identity matrices scaled by the appropriate variance because the levels within each random effects component are believed to be independent. When the levels are not independent, the identity matrix is replaced with a relationship matrix between the effects. For example, the additive (or numerator) relationship matrix (**A**) can be used to account for genetic relatedness between trees or families. The matrix can contain all the trees that have been measured across generations, as well as their unmeasured parents and relatives, assuming that the additive variance is the same for all of the populations that the origin parents come from (Henderson, 1984).

References

Bowyer, J. L. (2000). Wood science in a changing world—Where are we headed? *Wood Science and Technology*, *34*(3), 175–181. https://doi.org/10.1007/s002260000049

Cochran, W. G., & Cox, G. M. (1957). *Experimental designs*. Wiley.

Crossa, J. (2006). Modeling Genotype × Environment Interaction Using Additive Genetic Covariances of Relatives for Predicting Breeding Values of Wheat Genotypes (J. Crossa, J. Burgueño, P. L. Cornelius, G. McLaren, R. Trethowan, & A. Krishnamachari, Trans.). *Crop Science*, *v. 46*(4), 1722–1733. PubAg. https://doi.org/10.2135/cropsci2005.11-0427

Cullis, B., Gogel, B., Verbyla, A., & Thompson, R. (1998). Spatial Analysis of Multi-Environment Early Generation Variety Trials. *Biometrics*, *54*(1), 1–18. https://doi.org/10.2307/2533991

Dutkowski, G. W. (2005). *Improved models for the prediction of breeding values in trees* [University of Tasmania]. https://eprints.utas.edu.au/8835/1/Dutkowski_Thesis_Small.pdf

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.).

Henderson, C. R. (1984). *Applications of linear models in animal breeding*.

Isik, F., Goldfarb, B., LeBude, A., Li, B., & McKeand, S. (2005). Predicted genetic gains and testing efficiency from two loblolly pine clonal trials. *Canadian Journal of Forest Research*, *35*(7), 1754–1766. https://doi.org/10.1139/x05-064

Isik, F., Holland, J., & Maltecca, C. (2017). *Genetic Data Analysis for Plant and Animal Breeding*. Springer International Publishing. https://doi.org/10.1007/978-3-319-55177-7

Isik, F., & Mckeand, S. (2019). Fourth cycle breeding and testing strategy for Pinus taeda in the NC State University Cooperative Tree Improvement Program. *Tree Genetics & Genomes, 15*. https://doi.org/10.1007/s11295-019-1377-y

John, J. A., & Williams, E. R. (1995). *Cyclic and computer generated designs. Monographs on Statistics and Applied Probability 38.* Chapman & Hall.

Kelly, A. M. (2007). The Accuracy of Varietal Selection Using Factor Analytic Models for Multi-Environment Plant Breeding Trials (A. M. Kelly, A. B. Smith, J. A. Eccleston, & B. R. Cullis, Trans.). *Crop Science, v. 47*(3), 1063–1070. PubAg. https://doi.org/10.2135/cropsci2006.08.0540

Lauer, E., Sims, A., Mckeand, S., & Isik, F. (2021). Genetic Parameters and Genotype-by-Environment Interactions in Regional Progeny Tests of Pinus taeda L. in the Southern USA. *Forest Science*, *67*(1), 60–71. http://dx.doi.org/10.1093/forsci/fXaa035

Lindgren, D. (1991). Progeny testing. In *Developments in Plant Genetics and Breeding* (Vol. 3, pp. 191–203). Elsevier. https://doi.org/10.1016/B978-0-444-98724-2.50020-1

McKeand, S. E., Jokela, E., Huber, D., Byram, T., Allen, H., Li, B., & Mullin, T. (2006). Performance of improved genotypes of loblolly pine across different soils, climates, and silvicultural inputs. *Forest Ecology and Management*, 178–184. https://doi.org/10.1016/j.foreco.2006.02.016

McKeand, S. E. (2019). The Evolution of a Seedling Market for Genetically Improved Loblolly Pine in the Southern United States. Journal of Forestry, 117(3), 293–301. https://doi.org/10.1093/jofore/fvz006

McKeand, S. E., Payn, K. G., Heine, A. J., & Abt, R. C. (2021). Economic significance of continued improvement of loblolly pine genetics and its efficient deployment to landowners in the southern United States. *Journal of Forestry*, *119*(1), 62–72.

Mili, D., Kati, S., Bo, J., Miki, A., & Vasiljevi, S. (2010). *IMPORTANCE OF PROGENY TESTING IN ALFALFA BREEDING*. *42*(3), 8.

Mrode, R. A. (2014). *Linear models for the prediction of animal breeding values*. Cabi. https://books.google.com/books?hl=en&lr=&id=b6MKAwAAQBAJ&oi=fnd&pg=PR5&dq=mrode+linear+models&ots=8P4chw1ArS&sig=4a2yrp9UKJ3CDooOZUnoa3cdXiA

Ogut, F. (2012). *Predictions of Genetic Merit in Tree Breeding Using Factor Analytic Linear Mixed Models and Blended Genomic Relationship Matrices*. North Carolina State University.

Ogut, F., Maltecca, C., Whetten, R., McKeand, S., & Isik, F. (2014). Genetic Analysis of Diallel Progeny Test Data Using Factor Analytic Linear Mixed Models. *Forest Science*, *60*(1), 119–127. https://doi.org/10.5849/forsci.12-108

Patterson, H. D., & Williams, E. R. (1976). A new class of resolvable incomplete block designs. *Biometrika*, *63*(1), 83–92.

Patterson, H. D., Williams, E. R., & Hunter, E. A. (1978). Block designs for variety trials. *The Journal of Agricultural Science*, *90*(2), 395–400.

Piepho, H. P., Möhring, J., Melchinger, A. E., & Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, *161*(1), 209–228. https://doi.org/10.1007/s10681-007-9449-8

Piepho, H. P., & Williams, E. R. (2010). Linear variance models for plant breeding trials. *Plant Breeding*, *129*(1), 1–8. https://doi.org/10.1111/j.1439-0523.2009.01654.x

Sims, A. D. (2017). *Genetic Parameter Estimates from 3rd-Cycle Pollen Mix Progeny Tests in Loblolly Pine (Pinus taeda L.).* North Carolina State University.

Slamet, S. A. (2020). *Prediction of Breeding Values from Row-Column Designs in Pinus taeda Progeny Tests in the Southern US.* NCSU.

Smith, A.B., Cullis, B., & Thompson, R. (2001). Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. *Biometrics*, *57*(4), 1138–1147. https://doi.org/10.1111/j.0006-341X.2001.01138.x

Smith, A. B., Cullis, B. R., & Thompson, R. (2005). The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *The Journal of Agricultural Science*, *143*(6), 449–462. https://doi.org/10.1017/S0021859605005587

So, Y.-S. (2009). A Comparison of Mixed-Model Analyses of the Iowa Crop Performance Test for Corn (Y.-S. So & J. Edwards, Trans.). *Crop Science*, *v. 49*(5), 1593–0. PubAg. https://doi.org/10.2135/cropsci2008.09.0574

Spilke, J. (2005). Analysis of unbalanced data by mixed linear models using the MIXED procedure of the SAS system (J. Spilke, H. P. Piepho, & X. Hu, Trans.). *Journal of Agronomy and Crop Science*, *v. 191*(1), 47–54. PubAg. https://doi.org/10.1111/j.1439-037X.2004.00120.x

Turner, J. A., Buongiorno, J., & Zhu, S. (2006). An economic model of international wood supply, forest stock and forest area change. *Scandinavian Journal of Forest Research*, *21*(1), 73–86. https://doi.org/10.1080/02827580500478506

Whitaker, D., Williams, E. R., & John, J. A. (2001). CycDesigN: a package for the computer generation of experimental designs. *CSIRO Forestry and Forest Products: Canberra*.

White, L. (1996). Cyclic and Computer Generated Designs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *159*(3), 625–625. https://doi.org/10.2307/2983344

White, T. L., Adams, W. T., & Neale, D. B. (2007). *Forest Genetics*. CABI.

Williams, E. R., & Talbot, M. (1993). Alpha+, Experimental Designs for Variety Trials. *CSIRO, Canberra and SASS*.

Williams, E. R., & Matheson, A. C. (1994). *Experimental design and analysis for use in tree improvement*. CSIRO Information Services, East Melbourne, Australia.

Xiong, J. S., Isik, F., McKeand, S. E., & Whetten, R. W. (2010). *Genetic Variation of Stem Forking in Loblolly Pine*. 8.

Zobel, B., & Talbert, J. (1984). *Applied forest tree improvement.* John Wiley & Sons.

Chapter 2. Genetic Parameter Estimates for Growth Traits from *Pinus taeda* L. Fourth-Cycle Progeny Tests in the Coastal Breeding Population

2.1. Introduction

With the natural range spanning from East Texas to Maryland and New Jersey, loblolly pine (*Pinus taeda L.*) is the most economically valuable forest tree species in the southern United States (McKeand et al., 2004). The wide range of environmental diversity across the species' range has resulted in the separation of various geographic races (Wells, 1983; Zobel & Talbert, 1984; Schmidtling, 2001; Sierra-Lucero et al., 2002; McKeand et al., 2004). The sources from the Atlantic Coastal Plain are known for their rapid growth, poor stem shape, and lack of cold tolerance, whereas the Piedmont sources are slower growing with better stem form and cold tolerance (McKeand et al., 2004), and the Northern sources have relatively high rust disease resistance and tolerance to frost (Farjat et al., 2017).

The southern USA is the wood basket of the world. The region alone produces 17% of the world's wood and fiber by intensively managing pine plantations, especially *P. taeda* (Fox et al., 2007). In the region, improvement efforts on *P. taeda* have typically focused on increasing growth and disease resistance (White et al., 1993). The largest *P. taeda* breeding program is managed by the Cooperative Tree Improvement Program (CTIP) at North Carolina State University. The breeding efforts started in 1956, and the CTIP is currently implementing the fourth cycle of breeding, testing, and selection which started in 2014 (Isik & Mckeand, 2019). The major goal of the CTIP is to increase tree growth with testing, selection, and breeding. Because of a large

geographic distribution of the species and adaptation of land races in different regions, the CTIP split the breeding efforts into three major breeding zones, starting in the third cycle and continued in the fourth cycle (McKeand & Bridgwater, 1998; Isik & Mckeand, 2019).

Numerous publications showed that there is considerable genetic variation among seed sources and parents within seed sources for height and diameter growth (Li and Isik, 2003, Sims, 2017). In analyzing large number of diallels in the second cycle of *P. taeda* breeding, a wide range of additive, non-additive genetic variances and heritability estimates (from 0 to 1.0) were reported from 105 diallels (Isik et al., 2005). They reported an average heritability estimate of 0.19 for six-year height. In a recent study based on multiple series of multiple-environmental trials in the CTIP third cycle polymix tests, narrow-sense heritability estimates for tree height ranged from 0.18 to 0.24. while the estimates for DBH had a range of 0.11 to 0.21 (Lauer et al., 2021). Unlike 2nd-cycle six-parent diallels where parents were tested in narrow geographic regions, the third-cycle progeny tests had a larger number of half-sib families, and they were widely tested across geographic regions (Isik & Mckeand, 2019).

While estimates of genetic parameters are influenced by the degree of additive allelic control on traits, they are also influenced by the test site's quality (environmental factors) and data collecting (measurement errors) (Sims, 2017). In forestry, GxE can contribute to unpredictability in the performance of particular genotypes across a variety of conditions. An underlying goal has thus been to characterize the GxE in order to diminish or even eliminate such unpredictability (Li et al., 2017). The multi-environmental trials (MET) allow breeders to examine the amount of GxE interaction and genotype performance across several settings in addition to estimating genetic

parameters and breeding values (Li et al., 2017; Shalizi, 2020). CTIP tests loblolly pine in different environments across to the Southern US to estimate GxE effects. There are a remarkable number of studies of GxE in loblolly pine (McKeand et al. 2006; Ogut, 2012; Sims, 2017; Shalizi & Isik, 2019; Lauer et al., 2021) that examine GxE and the reasons behind it through different statistical approaches.

2.1.1 Objectives

The overall goal of this study is to estimate genetic parameters from the CTIP fourth-cycle progeny tests to draw conclusions about the magnitude of genetic variances, heritability and GxE. This study is somewhat different from previous works, because in the fourth cycle of *P. taeda* a different experimental design named alpha-cyclic incomplete row-column design (John & Williams, 1995) and rolling front progeny testing scheme has been implemented (Isik & Mckeand, 2019). Moreover, the test series over the years have much better connection compared to the diallels in the second cycle and polymix tests in the third cycle. Such considerable changes in progeny testing in the fourth cycle are expected to produce genetic parameters that are different from previously published work. The objectives of this study were 1) to estimate genetic parameters such as additive and non-additive genetic variance components and heritability, 2) calculate narrow-sense and broad-sense individual heritabilities, and 3) examine GxE interaction for 4^{th}-Cycle coastal progeny test data of loblolly pine for growth traits (height and DBH).

2.2. Material and Methods

2.2.1 Genetic material

The North Carolina State University (NCSU) Cooperative Tree Improvement Program (CTIP) started the 4th-Cycle breeding strategy in 2014 (Isik & Mckeand, 2019). One of the main goals of the breeding strategy was to evaluate 800 crosses from the Atlantic Coastal breeding population. Due to large number of crosses planned, a rolling-front mating design and progeny testing strategy was implemented. Every year a new set of crosses were produced. The mating design was carried out using the MateSel software (Kinghorn, 2011). For each parent tree, the MateSel software calculated an index breeding value by combining stem volume, stem straightness, and fusiform rust disease incidence. For brevity, we will give details about the Atlantic Coastal breeding population since it is the largest population and are the data used in this study.

In the Coastal breeding population, 421 unique parents selected from the third cycle were used in the mating design. For each parent tree, an index breeding value was calculated by combining stem volume (60%), stem straightness (20%), and fusiform rust disease incidence (20%). The MateSel software produced 746 crosses each with about 60 full-sib progenies tested in multi environmental trials across the Coastal Plains of southern US (Isik & Mckeand, 2019).

In this study, data from 25 progeny tests established between 2014 to 2017 were utilized. The data had 376 full-sib families generated from 294 unique parents. The total number of progeny assessed was approximately 17,800. The remaining progeny tests had not yet been assessed.

2.2.2 Experimental design

Seedling progeny tests used in this research were established in 25 locations in the Southeastern US (**Figure 2.1**). The experimental design was alpha-cyclic incomplete row-column design (John & Williams, 1995; Williams et al., 2006). This computer-generated design is efficient in controlling environmental noise in row-column directions (John & Williams, 1995). The rows and columns are regarded as incomplete blocks increasing analytical power in linear mixed models.

Seedlings from each cross were assigned to specific rows and columns to optimize testing efficiency (reducing the residual variance and increasing concurrence) and to improve direct comparisons among crosses. Seedling location in one rep is linked to seedling location in all reps throughout a testing series for a given year, both at the same site and at other sites. Families were typically tested across four years between 2014 and 2017. To provide connection and comparison of tests across years, common families were included in the testing strategy across multiple years. The 4^{th}-Cycle Coastal progeny test data used in this research were comprised of 294 unique parents. Test sites and test series were strongly connected with each other. The number of shared parents ranged between 38 and 164 among site pairs. Test sites within a planting year were strongly connected compared with the test sites among years. Tests planted in 2014 had the lowest number of shared parents (38 – 41) with the tests planted in 2017 (**Figure 2.2**).

2.2.3 Data collection

Depending on the growth, tests were assessed at ages four or five years. Most tests were assessed at age four. Tree height (m), diameter at breast height (DBH, cm), incidence of fusiform rust diseases (yes / no), and incidence of forking and ramicorn were recorded. For this chapter, we analyzed height and DBH for the data set consisting of 376 full-sib families, 294 unique parents,

and approximately 17800 individual trees assessed through 2014 to 2017 field trials (**Figure 2.1**). R statistical software was used for data cleaning, visualization, and producing descriptive and summary statistics (R development Core Team, 2021).

2.2.4 Statistical analysis

Height and DBH were analyzed using the following linear mixed model in multi-environment trial (MET) setting to estimate variance components, and genetic merit of 294 parents of the 4th-Cycle Coastal population.

$$y = \mu + \boldsymbol{X}s + \boldsymbol{Z}_1 b + \boldsymbol{Z}_2 q + \boldsymbol{Z}_3 c + \boldsymbol{Z}_4 a + \boldsymbol{Z}_5 f + \boldsymbol{Z}_6 i + e \quad (2.1)$$

Where y is the vector of response variable; μ is overall mean; $\boldsymbol{X}$ and $\boldsymbol{Z}$ are incidence matrices for the fixed and random effects, respectively; s is the vector of site fixed effect ($s = 25$); b is the vector for random replicate effect nested within site with $b \sim MVN(0, \mathrm{I}\sigma_b^2)$; q is the vector of random row effect nested within rep with $q \sim MVN(0, I\sigma_q^2)$; c is the vector of random column effect nested within rep with $c \sim MVN(0, I\sigma_c^2)$; a is the vector of random tree id effect nested within test site with $a \sim MVN(0, \boldsymbol{G}_a \otimes \boldsymbol{A})$; f is the vector of random family genetic effects with $f \sim MVN(0, \boldsymbol{I}\sigma_f^2)$; i is the vector of random family by site effect with $i \sim MVN(0, \boldsymbol{I}\sigma_i^2)$; and e is the vector of random errors with $e \sim MVN(0, \boldsymbol{R})$. Where, $\boldsymbol{I}$ is an identity matrix of its proper dimensions, $\boldsymbol{G}_a$ is the s-by-s variance covariance matrix of tree id nested within site effect, $\boldsymbol{A}$ is the numerator relationship matrix calculated from pedigree, and the symbol $\otimes$ is direct product operator.

The assumptions for the random replicate ($\boldsymbol{B}$), row ($\boldsymbol{Q}$), column ($\boldsymbol{C}$), and residual ($\boldsymbol{R}$) effects were relaxed with block diagonal (BD) variance structure. This structure provided independent units for each site ($\bigoplus_{i=1}^{S} \sigma_i^2 \boldsymbol{I}_{n_i}$). The random family ($f$) and family-by-site ($i$) effects were fit with independent and identically distributed (IID) variance structure ($\boldsymbol{I}\sigma^2$). The additive genetic effects ($\boldsymbol{G}_a$) part of the model was fit with IID, heterogeneous (CORUH), and extended factor analytic of order one (XFA1) variance-covariance structures. The mathematical form of IID $\boldsymbol{G}_a$ variance structure is (Isik et al., 2017);

$$\boldsymbol{G}_a \otimes \boldsymbol{A} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & \vdots \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2 \end{bmatrix} \otimes \boldsymbol{A} \tag{2.2}$$

where, σ^2 is the uniform additive genetic variance across all sites and 0 values in the off-diagonal of the matrix show no covariance between pairs of sites, and **A** is the numerator relationship matrix derived from the pedigree. The mathematical form of CORUH $\boldsymbol{G}_a$ variance structure is (Isik et al., 2017);

$$\boldsymbol{G_a} \otimes \boldsymbol{A} = \begin{bmatrix} \sigma_{a1}^2 & \sigma_{a1}\sigma_{a2}\rho & \cdots & \sigma_{a1}\sigma_{a25}\rho \\ \sigma_{a1}\sigma_{a2}\rho & \sigma_{a2}^2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \sigma_{a24}\sigma_{a25}\rho \\ \sigma_{a1}\sigma_{a25}\rho & \cdots & \sigma_{a24}\sigma_{a25}\rho & \sigma_{a25}^2 \end{bmatrix} \otimes \boldsymbol{A} \tag{2.3}$$

where, the diagonal elements are site specific additive genetic variances and off-diagonal elements are uniform covariance between pairs of sites. This covariance structure provided a uniform correlation among site pairs.

To better evaluate additive genotype-by-environment interactions, a multiplicative factor analytic of order *k* = 1 was also fit to model the random $\boldsymbol{G_a}$ covariance structure. A factor analytic structure regresses genetic and GxE effects on *k* unknown environmental covariates (Smith et al., 2001). This structure provided unique additive genetic variance for each site and unique additive genetic correlation between pairs of sites. The mathematical form of this structure can be written as (Smith et al., 2001).

$$\boldsymbol{G_a} \otimes \boldsymbol{A} = [\, \boldsymbol{\Lambda\Lambda}^T + \boldsymbol{\psi} \,] \otimes \boldsymbol{A} = \begin{bmatrix} \lambda_{11}^2 + \psi_1 & \lambda_{11}\lambda_{12} & \cdots & \lambda_{11}\lambda_{125} \\ \lambda_{11}\lambda_{12} & \lambda_{12}^2 + \psi_2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \cdots \\ \lambda_{11}\lambda_{125} & \cdots & \cdots & \lambda_{125}^2 + \psi_{25} \end{bmatrix} \otimes \boldsymbol{A} \tag{2.4}$$

where, $\boldsymbol{\Lambda}$ is the *s*-by-*k* (*s* = site, *k* = number of factors) matrix of site loadings and $\boldsymbol{\psi}$ is the *s*-by-*s* matrix of site-specific variances, λ_{11}^2 is the variance explained by site loading of factor 1 at site 1, ψ is the site-specific variance, and off-diagonals are correlations. Models were compared for the best $\boldsymbol{G_a}$ fit using residual log likelihood (LogL) and Akaike Information Criterion (AIC).

Individual tree narrow-sense (h_i^2) and broad-sense (H_i^2) heritabilities were estimated for height and DBH using the output from XFA1 model which was the best fit among all models. Heritabilities were estimated using the equations from (Isik et al., 2017).

$$h_i^2 = \frac{\bar{\sigma}_a^2\, \bar{r}_a}{\bar{\sigma}_a^2 + \sigma_f^2 + \sigma_i^2 + \bar{\sigma}_e^2} \tag{2.5}$$

$$H_i^2 = \frac{\bar{\sigma}_a^2\, \bar{r}_a + 4\sigma_f^2}{\bar{\sigma}_a^2 + \sigma_f^2 + \sigma_i^2 + \bar{\sigma}_e^2} \tag{2.6}$$

where, $\bar{\sigma}_a^2$ is the tree variance averaged across 25 sites; $\bar{r}_a$ is the pairwise additive genetic by site correlation averaged across 25 sites; σ_f^2 is the cross variance; σ_i^2 is the cross by site interaction variance; $4\sigma_f^2$ is the dominance genetic variance; $\bar{\sigma}_e^2$ is the residual variance averaged across 25 sites. Standard error of heritabilities were estimated using delta method (Lynch & Walsh, 1998).

To evaluate GxE interactions, the additive genetic correlation matrix from XFA1 model was first converted into a distance matrix and then clustered using dissimilarity agglomerative hierarchical clustering analysis. The cluster analysis was performed using the *hclust* function to create the dendrogram in R statistical software (R development Core Team, 2021). A heatmap with a dendrogram was created using the *heatmap.2* function of the *gplots* package in R statistical software. The full-sib family effect was fit with simple IID variance structure. Thus, we used the following equation to calculate dominance genetic correlation across all sites.

$$r_D = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_i^2} \tag{2.7}$$

The terms in the equation were explained previously.

The following multiple linear regression analysis was performed to explore effects of difference in mean minimum winter temperature (ΔMWT) and difference in spatial distance (S) among site pairs on additive GxE correlations.

$$y = \beta_0 + \beta_1 \Delta S + \beta_2 \Delta MWT + \beta_3 \Delta S * \Delta MWT + \varepsilon \tag{2.8}$$

where, y is the response variable additive genetic correlation between site pairs, β_0 is the intercept, β_1 is the slope for difference in spatial distance (ΔS) among site pairs, β_2 is the slope for difference in MWT among site pairs, β_3 is the slope for interaction between ΔS and ΔMWT, and ε is the random error with $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$. The difference in MWT (°C) was calculated as $\Delta MWT = |MWT_s - MWT_{s'}|$. The spatial distance was obtained by taking the sum of absolute difference in latitude and longitude among site pairs as $\Delta S = |Lat_s - Lat_{s'}| + |Long_s - Long_{s'}|$.

2.3. Results

2.3.1. Summary statistics

Height and DBH summary statistics for test sites are provided in **Figure 2.3** and **Figure 2.4**, respectively. Detailed summary statistics for the two traits are also provided in **Table A1** (height) and **Table A2 (DBH)** in the Appendix A. Test sites showed large variation for height and DBH growth at ages four and five years. Height growth ranged between 3.1 m and 6.7 m among sites with an average of 5.2 m across test sites. DBH ranged between 4.9 cm and 11.5 cm among sites, with an average 8.8 cm across test sites.

2.3.2. Model selection

Extended factor analytic of order one (XFA1) variance-covariance structure was the best fit for additive genetic ($\boldsymbol{G_a}$) effects compared with independent and identically distributed (IID) and heterogeneous (CORUH) structures for both height and DBH (**Table 2.1**). Model fit statistics, LogL and AIC improved substantially when $\boldsymbol{G_a}$ variance-covariance was fit with the XFA1. The first factor in the XFA1 model explained 70 to 73 % of additive genetic variance for DBH and height, respectively. We did not fit factor analytic of order two or higher due to convergence issues. The best model XFA1 was used to estimate variance components and explore GxE interaction for height and DBH in the *P. taeda* 4th-Cycle Coastal population.

2.3.3. Variance components and heritability

Genetic parameter estimates for height and DBH calculated from the XFA1 model are presented in **Table 2.2**. Additive genetic variance explained 56 and 67% of total genetic variation in DBH and height, respectively. Additive genetic variance was 1.3 times greater than the

dominance variance in DBH. The ratio between additive and dominance genetic variance was 2.1 for height. A large proportion of the phenotypic variance was residual variance ranging from 76 and 82% for the two traits.

Individual-tree narrow-sense and broad-sense heritability estimates were larger for height compared with the heritabilities for DBH. For height, individual tree narrow-sense and broad-sense heritabilities were 0.15 and 0.22, respectively. The two heritabilities were 0.09 and 0.17 for DBH. All heritability estimates were associated with small standard errors (**Table 2.2**).

2.3.4. GxE interactions

Additive and dominance GxE interactions for height and DBH were assessed using the GxE correlations calculated from the linear mixed models. Additive genetic correlations were 0.69 and 0.56 for height and DBH. Dominance genetic correlations were high for both height (0.84) and DBH (0.81) compared with the additive genetic correlations indicating negligible dominance GxE interactions among all site pairs (**Table 2.2**).

The XFA1 $\boldsymbol{G_a}$ structure provided additive genetic correlations for all site pairs. In total, 300 unique pairwise additive genetic correlations were calculated. The pairwise additive GxE correlations are presented in form of a heatmap in **Figure 2.5** and **Figure 2.6** for height and DBH, respectively. For height, the additive GxE correlation averaged 0.69, ranged 0.10 to 0.99 (**Table 2.2, Figure 2.5**). Whereas, for DBH, the additive GxE correlation averaged 0.56 and ranged between -0.44 and 0.99 (**Table 2.2, Figures 2.6**).

The heatmap and dendrogram clustered test sites with similar additive GxE correlations. For height, the dendrogram clustered the Coastal 4th-Cycle test sites into three major categories.

The first category included 18 test sites showing negligible additive GxE interactions (>0.60), depicted as dark blue color in **Figure 2.5**. The second category was comprised of five test sites showing low to moderate additive GxE interactions (~0.30 to ~0.60) with all test sites. The third category included two test sites (2016PCL02 and 201WEY06) expressing high additive GxE interactions (0.10 to ~0.30) with all sites (**Figure 2.5**).

The additive GxE interactions were much greater for DBH compared with height. Similar to height, the DBH additive GxE correlation heatmap was clustered into three categories (**Figure 2.6**). The first cluster was comprised of 18 test sites (dark blue color in the heatmap) showing moderate to low additive GxE interactions (>0.50) with each other. The second cluster included three test sites showing strong additive GxE interactions (0.10 to ~0.40) with all test sites. The third category of sites showed negative additive GxE correlations ranging from -0.44 to 0 (**Figure 2.6**).

The linear regression analysis of additive GxE correlations (y-axis) indicated a significant but small relationship with the difference in spatial distance among site pairs (**Table 2.3**). In general, test sites that were distantly apart in latitude and longitude showed higher additive GxE interactions than the test sites located near each other (**Figure 2.7**). The difference in MWT did not affect GxE interactions, and its correlation with the response variable was essentially zero for both traits.

2.4. Discussion

2.4.1. Model selection

In this study, the 4th-Cycle Coastal breeding population was evaluated in 25 seedling progeny tests established across the southeastern US. Due to large number of test trials, complex variance-covariance structures were fit to model additive genetic effects in MET linear mixed models. Factor analytic of order one was the best fit and most parsimonious compared to models fit with IID and heterogenous covariance structures. These findings support results of previously published research comparing factor analytic models with other variance-covariance structures such as IID, compound symmetry, CORUH, and unstructured in MET analysis of *P. taeda* (Zapata-Valenzuela, 2012; Gezan et al., 2017; Shalizi & Isik, 2019; Lauer et al., 2021). Additionally, use of factor analytic covariance structure for the additive genetic effects allowed for exploration of additive GxE interactions among the 25 sites which could not be possible with simpler variance-covariance structures.

2.4.2. Variance components and heritability

In this study, additive genetic variance explained a large proportion of the total genetic variance for DBH and height suggesting that the two traits are largely controlled by many genes with small effects. However, dominance variance was also significant, explaining 44 and 32% of genetic variation for the two traits, respectively. Numerous studies on *P. taeda* cloned and seedling progeny data reported higher additive genetic variance compared to dominance or non-additive (dominance + epistasis) genetic variance (Li & Isik, 2003; Isik et al., 2005; McKeand et al., 2008; Cumbie et al., 2012; de Almeida Filho et al., 2016; Shalizi & Isik, 2019). In those studies, the ratio

between additive and dominance variance ranged between 0.9 to 2.1 for growth traits (height, DBH, and volume), corresponding with the results observed in this study.

In this study, significant dominance variance in height and DBH could be due to the large number of full-sib families (376) used which helped better partition variance components into additive and dominance variances. However, high dominance variance estimates may also be related to strong genetic connectedness because the parents used in the mating design originated from a small breeding population. Published studies on *P. taeda* linked higher dominance variance with the mating of closely related parents (Isik, 2005; Ogut, 2012). In a study based on 105 diallels in the second cycle of *P. taeda* breeding program, non-additive genetic variance explained a small fraction of the total genetic variance (Isik et al., 2005).

Individual-tree heritability (narrow-sense and broad-sense) estimates were higher for height compared to DBH, suggesting stronger genetic control of the trait (**Table 2.2**). Lower heritability estimates for DBH could be associated with higher dominance and residual variances observed for the trait. The two parameters explained 8 and 82% of the phenotypic variance, respectively, while additive genetic effects explained only 10% of the phenotypic variation. Additionally, the additive GxE correlations were low (averaged 0.56) resulting in smaller additive genetic variance for DBH in this population. The heritability estimates of height and DBH in this study are comparable with the heritability estimates reported in numerous published studies on *P. taeda* half-sib and full-sib seedling progeny data (Isik et al., 2005; Cumbie et al., 2012; Ogut, 2012; Farjat, 2015; McKeand et al., 2008). In those studies, individual-tree narrow-sense heritability of growth traits ranged between 0.13 and 0.62.

2.4.3. GxE interpretation

The MET analysis of the 4th-Cycle Coastal progeny test data revealed significant additive GxE interactions among some of the site pairs for both height and DBH. In both traits, a group of test sites clustered together showing high site-to-site correlations (**Figures 2.5 and 2.6**). Two sites (2015WEY06 and 2016PLC02) showed significant additive GxE interactions with other sites. Similarly, for DBH, sites (2015WEY05 and 2016PLC02) had low- additive site-to-site correlations (significant GxE interactions). The site-site additive GxE interactions were more pronounced for DBH compared to height suggesting some rank change of genotypes across environments. The reasons for such significant GxE between above sites and the rest of the sites in unknown.

In this study, several environmental factors could be related to additive GxE interactions. The linear regression analysis suggested that the difference in spatial distance among site pairs is important to explain some of the additive GxE interactions. As expected, site that were distantly apart from each other on geographic landscape showed higher additive GxE interactions compared with the sites located near each other, but the association was not strong. This was very true for the WEY test sites mentioned above because they were distantly apart from other sites. Published studies on *P. taeda* seedling progeny test data report similar results in the southeastern US (Farjat et al., 2017; Lauer et al., 2021). Relatively lower site-to-site correlations between pairs of sites in the fourth cycle tests should be interpreted cautiously because unlike the third cycle polymix tests (McKeand & Bridgwater, 1998), the fourth cycle tests are not balanced (Isik & McKeand, 2019). The GxE statistics are more reliable when the test sites are balanced.

The difference in minimum winter temperature between test sites was correlated (0.54) with the difference in spatial distance, but the difference in minimum winter temperature among site pairs did not affect GxE interactions which was also confirmed by Lauer et al. (2021). In both studies, the Coastal progeny tests were established on environments within the minimum winter temperature guidelines recommended for the *P. taeda* seed movement (Schmidtling, 2001). As a result, the effects of difference in minimum winter temperature among site pairs was not evident, because the trees were not tested on extreme sites.

In factor analytic MET modeling, it is hard to identify unknown environmental covariates related to GxE. As discussed previously, spatial distance could be the unknown environmental covariate for the first factor. However, other environmental factors such as fusiform rust disease incidence, site index, tip moth, precipitation, and frost incidence could be also linked to additive GxE interactions (Costa e Silva et al., 2006; Chen et al., 2017; Ukrainetz et al., 2018; Shalizi & Isik, 2019).

2.5. Conclusions

Factor analytic genetic covariance structure was the best fit to model additive GxE interactions compared with the simpler covariance structures for growth traits in the multi-environmental progeny tests of *P. taeda* Coastal 4th-Cycle Breeding Population. The narrow-sense heritability estimate for height (0.15) was somewhat lower in this study compared to published results, whereas the estimate for DBH was similar to previously published results. The overall additive genetic correlations for height (0.69) and for DBH (0.56) were considerably lower than published results. Two sites contributed significantly to these low estimates. The rest of 25 sites had high additive site-to-site correlations, suggesting the non-significant GxE in most of the 4th-Cycle Coastal breeding population.

2.6. References

Chen, Z.-Q., Karlsson, B., & Wu, H. X. (2017). Patterns of additive genotype-by-environment interaction in tree height of Norway spruce in southern and central Sweden. *Tree Genetics & Genomes*, *13*(1), 25. https://doi.org/10.1007/s11295-017-1103-6

Costa e Silva, J., Potts, B. M., & Dutkowski, G. W. (2006). Genotype by environment interaction for growth of Eucalyptus globulus in Australia. *Tree Genetics & Genomes*, *2*(2), 61–75. https://doi.org/10.1007/s11295-005-0025-x

Cumbie, W. P., Isik, F., & McKeand, S. E. (2012). Genetic Improvement of Sawtimber Potential in Loblolly Pine. *Forest Science*, *58*(2), 168–177. https://doi.org/10.5849/forsci.09-060

de Almeida Filho, J. E., Guimarães, J. F. R., e Silva, F. F., de Resende, M. D. V., Muñoz, P., Kirst, M., & Resende, M. F. R. (2016). The contribution of dominance to phenotype prediction in a pine breeding and simulated population. *Heredity*, *117*(1), 33–41. https://doi.org/10.1038/hdy.2016.23

Farjat, A. E. (2015). *Optimal Seed Deployment under Climate Change using Spatial Models and Prediction of Genetic Merit in Loblolly Pine. (Under the direction of Brian J. Reich and Fikret Isik.)*. NCSU.

Farjat, A. E., Chamblee, A. K., Isik, F., Whetten, R. W., & McKeand, S. E. (2017). Variation among Loblolly Pine Seed Sources across Diverse Environments in the Southeastern United States. *Forest Science*, *63*(1), 39–48. http://dx.doi.org/10.5849/forsci.15-107

Fox, T. R., Allen, H. L., Albaugh, T. J., Rubilar, R., & Carlson, C. A. (2007). Tree Nutrition and Forest Fertilization of Pine Plantations in the Southern United States. *Southern Journal of Applied Forestry*, *31*(1), 5–11.

Gezan, S. A., de Carvalho, M. P., & Sherrill, J. (2017). Statistical methods to explore genotype-by-environment interaction for loblolly pine clonal trials. *Tree Genetics & Genomes*, *13*(1), 1. https://doi.org/10.1007/s11295-016-1081-0

Isik, F., Boos, D. D., & Li, B. (2005). The distribution of genetic parameter estimates and confidence intervals from small disconnected diallels. *Theoretical and Applied Genetics*, *110*(7), 1236–1243. https://doi.org/10.1007/s00122-005-1957-0

Isik, F., Goldfarb, B., LeBude, A., Li, B., & McKeand, S. (2005). Predicted genetic gains and testing efficiency from two loblolly pine clonal trials. *Canadian Journal of Forest Research*, *35*(7), 1754–1766. https://doi.org/10.1139/x05-064

Isik, F., Holland, J., & Maltecca, C. (2017). *Genetic Data Analysis for Plant and Animal Breeding*. Springer International Publishing. https://doi.org/10.1007/978-3-319-55177-7

Isik, F., & McKeand, S. E. (2019). Fourth cycle breeding and testing strategy for Pinus taeda in the NC State University Cooperative Tree Improvement Program. *Tree Genetics & Genomes*, *15*(5), 70. https://doi.org/10.1007/s11295-019-1377-y

John, J. A., & Williams, E. R. (1995). *Cyclic and computer generated designs. Monographs on Statistics and Applied Probability 38.* Chapman & Hall.

Kinghorn, B. P. (2011). An algorithm for efficient constrained mate selection. *Genetics Selection Evolution*, *43*(1), 4. https://doi.org/10.1186/1297-9686-43-4

Lauer, E., Sims, A., Mckeand, S., & Isik, F. (2021). Genetic Parameters and Genotype-by-Environment Interactions in Regional Progeny Tests of Pinus taeda L. in the Southern USA. *Forest Science*, *67*(1), 60–71. http://dx.doi.org/10.1093/forsci/fXaa035

Li, B., & Isik, F. (2003). Time Trend of Genetic Parameters in Growth Traits of Pinus taeda L. *Silvae Genetica*.

Li, Y., Suontama, M., Burdon, R. D., & Dungey, H. S. (2017). Genotype by environment interactions in forest tree breeding: Review of methodology and perspectives on research and application. *Tree Genetics & Genomes*, *13*(3), 60. https://doi.org/10.1007/s11295-017-1144-x

Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates Incorporated. https://www.journals.uchicago.edu/doi/10.1086/393101

McKeand, S. E., & Bridgwater, F. E. (1998). A strategy for the third breeding cycle of loblolly pine in the southeastern US. *Silvae Genetica*, *47*(4), 223–234.

McKeand, S. E., Kegley, A. J., & Li, B. (2004). Seedling Evaluation of Atlantic Coastal and Piedmont Sources of Loblolly Pine and Their Hybrids for Height Growth. *Southern Journal of Applied Forestry*, *28*(2), 83–90. https://doi.org/10.1093/sjaf/28.2.83

McKeand, S. E., Jokela, E., Huber, D., Byram, T., Allen, H., Li, B., & Mullin, T. (2006). Performance of improved genotypes of loblolly pine across different soils, climates, and silvicultural inputs. *Forest Ecology and Management*, 178-184. https://doi.org/10.1516/j.foreco.2006.02.016

McKeand, S. E., Li, B., Grissom, J. E., Isik, F., & Jayawickrama, K. J. S. (2008). Genetic Parameter Estimates for Growth Traits from Diallel Tests of Loblolly Pine Throughout the Southeastern United States. *Silvae Genetica*, *57*(1–6), 101–110. https://doi.org/10.1515/sg-2008-0016

Ogut, F. (2012). *Predictions of Genetic Merit in Tree Breeding Using Factor Analytic Linear Mixed Models and Blended Genomic Relationship Matrices*. North Carolina State University.

R development Core Team. (2021). *R: A language and environment for statistical computing* (4.1.1) [Computer software]. R Foundation for Statistical Computing.

Schmidtling, R. C. (2001). *Southern Pine Seed Sources*. *44*, 35.

Shalizi, M. N., & Isik, F. (2019). Genetic parameter estimates and GxE interaction in a large cloned population of Pinus taeda L. *Tree Genetics & Genomes*, *15*(3), 46. https://doi.org/10.1007/s11295-019-1352-7

Shalizi, M. N. (2020). *Genetic Parameter Estimates and Prediction of Genetic Merit in Clonal Populations of Pinus taeda L. using Expected and Realized Genomic Relationships* [NCSU].https://repository.lib.ncsu.edu/bitstream/handle/1840.20/38351/etd.pdf?sequence=1&isAllowed=y

Sierra-Lucero, V., McKeand, S. E., Huber, D. A., Rockwood, D. L., & White, T. L. (2002). Performance differences and genetic parameters for four coastal provenances of Loblolly Pine in the southeastern United States. *Forest Science*, *48*(4), 732–742.

Sims, A. D. (2017). *Genetic Parameter Estimates from 3rd-Cycle Pollen Mix Progeny Tests in Loblolly Pine (Pinus taeda L.).* North Carolina State University.

Smith, A., Cullis, B., & Thompson, R. (2001). Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. *Biometrics*, *57*(4), 1138–1147. https://doi.org/10.1111/j.0006-341X.2001.01138.x

Ukrainetz, N. K., Yanchuk, A. D., & Mansfield, S. D. (2018). Climatic drivers of genotype–environment interactions in lodgepole pine based on multi-environment trial data and a factor analytic model of additive covariance. *Canadian Journal of Forest Research*, *48*(7), 835–854. https://doi.org/10.1139/cjfr-2017-0367

Wells, O. O. (1983). Southwide Pine Seed Source Study—Loblolly Pine at 25 Years. *Southern Journal of Applied Forestry, 7*(2), 63–71. https://doi.org/10.1093/sjaf/7.2.63

White, T. L., Hodge, G. R., & Powell, G. L. (1993). An advanced-generation tree improvement plan for slash pine in the southeastern United States. *Silvae Genetica, 42*(6), 359–371.

Williams, E. R., John, J. A., & Whitaker, D. (2006). Construction of Resolvable Spatial Row–Column Designs. *Biometrics*, *62*(1), 103–108. https://doi.org/10.1111/j.1541-0420.2005.00393.x

Zapata-Valenzuela, J. (2012). Use of Analytic Factor Structure to Increase Heritability of Clonal Progeny Tests of Pinus taeda L. *Chilean Journal of Agricultural Research, 72*(3), 309–315. https://doi.org/10.4067/S0718-58392012000300002

Zobel, B., & Talbert, J. (1984). *Applied forest tree improvement.* John Wiley & Sons.

Table 2.1 GxE variance-covariance structure, model fit statistics (log likelihood, Akaike Information Criterion (AIC)), and number of parameters for the genetic structure for height and DBH. Factor analytic covariance structure produced the best AIC values.

Trait	**G_a structure**	**LogL**	**AIC**	**G_a parameters (total)**
Height	IID - $(\sigma_g^2 J_s + \sigma_{ge}^2 I_s) \otimes A$	-1758	3724	2 (104)
	CORUH - $(\Sigma \otimes A)$	-1715	3686	26 (128)
	XFA1 - $\Sigma(\Lambda\Lambda^T + \psi) \otimes A$	-1691	3673	50 (152)
DBH	IID - $(\sigma_g^2 J_s + \sigma_{ge}^2 I_s) \otimes A$	-4892	29992	2 (104)
	CORUH - $(\Sigma \otimes A)$	-4843	29943	26 (128)
	XFA1 - $\Sigma(\Lambda\Lambda^T + \psi) \otimes A$	-4815	29917	50 (152)

Table 2.2 Additive genetic variance (σ_a^2), non-additive genetic variance (σ_{na}^2), the ratio between two genetic variances (a/na), residual variance (σ_e^2), narrow-sense individual heritability (h^2), broad-sense individual heritability (H^2), average additive genetic correlations (range) between pairs of sites (r_A), and average dominance genetic correlations (range) between pairs of sites (r_D) of height and DBH from XFA1 models.

Estimates	**Height ±SE**	**DBH ±SE**
Additive variance (σ_a^2)	0.077 ±0.001	0.228 ±0.0310
Dominance variance (σ_d^2)	0.036 ±0.009	0.178 ±0.0110
σ_a^2/σ_d^2	2.1	1.3
Error variance (σ_e^2)	0.380 ±0.010	1.908 ±0.044
Narrow-sense heritability (h_i^2)	0.15 ±0.018	0.09 ±0.012
Broad-sense heritability (H_i^2)	0.22 ±0.021	0.17 ±0.018
Additive genetic correlation (r_A)	0.69 ±0.015	0.56 ±0.016
Dominance genetic correlation (r_D)	0.84 ±0.070	0.81 ±0.060

Table 2.3 Analysis of variance results from the multiple regression using additive GxE correlations for tree height and DBH as the response variable and the difference in spatial distance and difference in mean minimum winter temperature (°C) as explanatory variables. The difference in spatial distance explained a significant amount of variation for both height and DBH.

	Model term	**Mean square**	**F-value**	**P-value**
Height	Difference in spatial distance (ΔS)	1.01	18.98	**<0.001**
	Difference in MWT (ΔMWT)	0.02	0.32	0.57
	Interaction $\Delta S * \Delta MWT$	0.12	2.23	0.14
	Residuals	0.05		
DBH	Difference in spatial distance (ΔS)	1.74	10.45	**<0.001**
	Difference in MWT (ΔMWT)	0.61	3.66	0.06
	Interaction $\Delta S * \Delta MWT$	0.39	2.36	0.13
	Residuals	0.17		

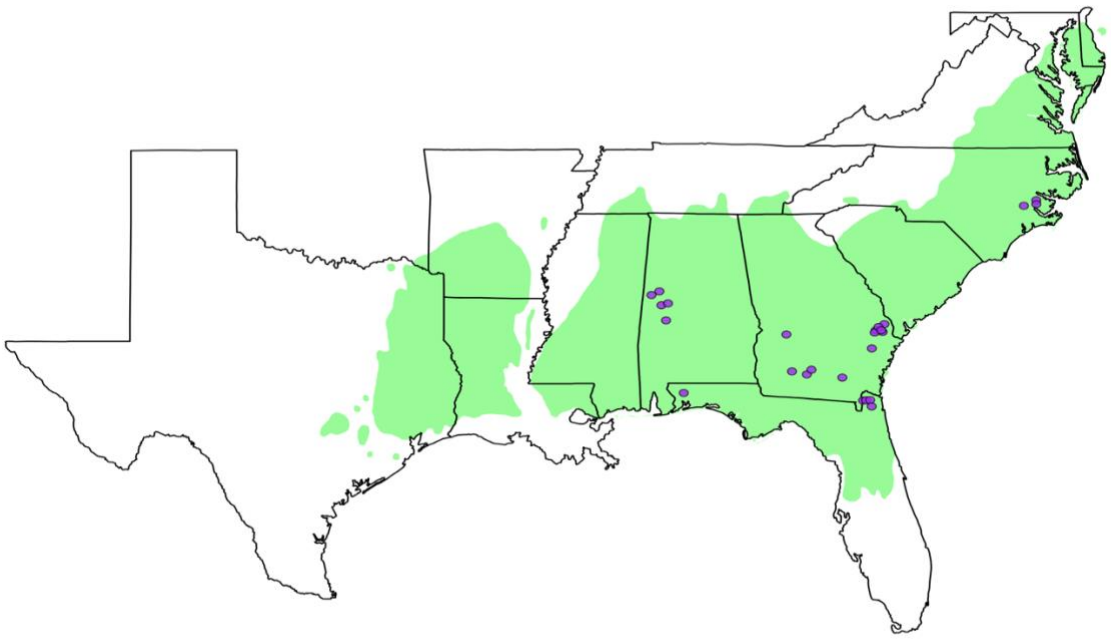

Figure 2.1 Map of the geographic locations of the test sites in the Coastal breeding population. Purple dots are test sites of the 4th-cycle Coastal *Pinus taeda* breeding population used in this research while shaded area is the natural range of *Pinus taeda*.

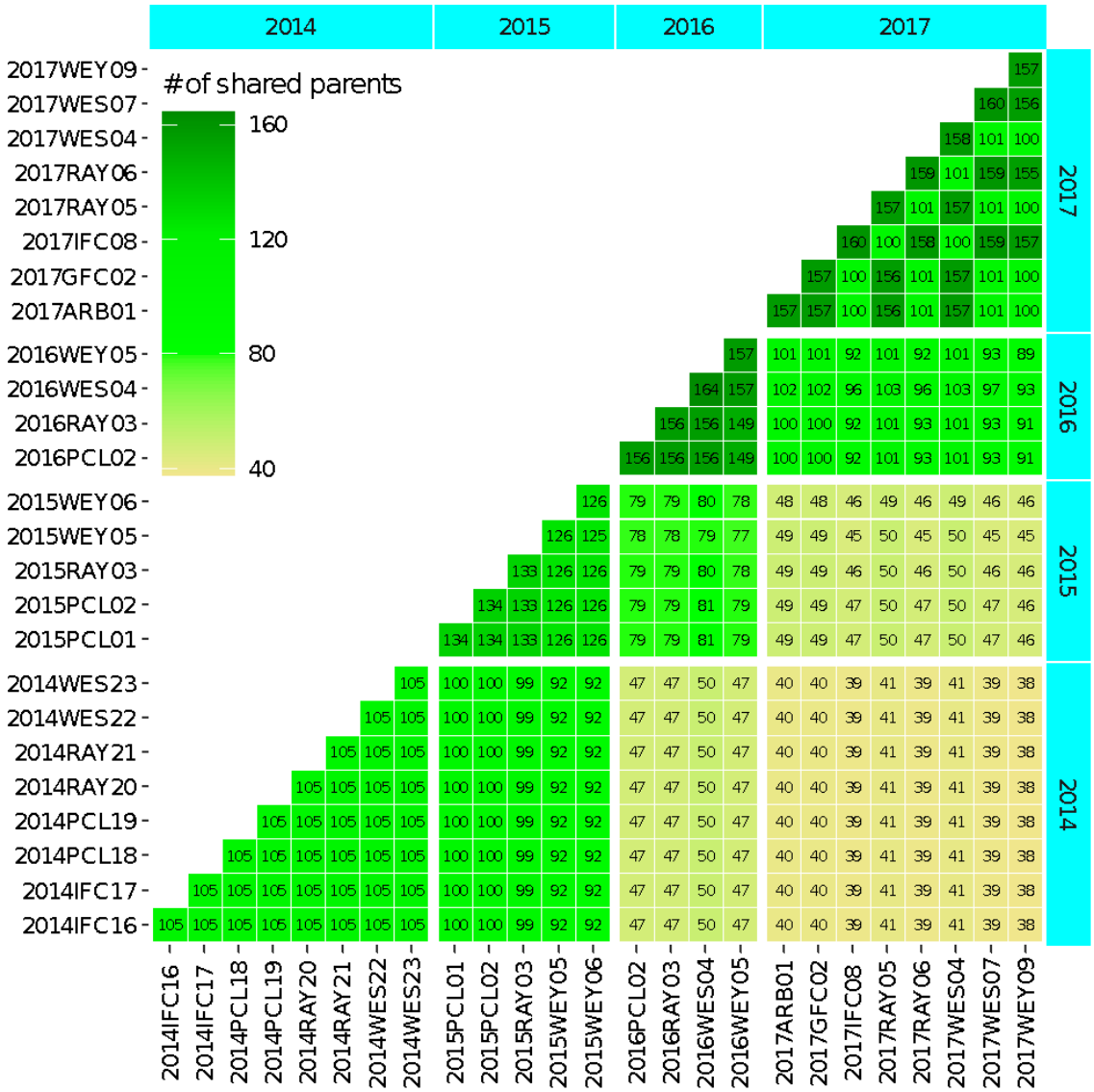

Figure 2.2 Connection matrix shows how many parents (female or male) are shared between each test sites. The weakest connection was between the first test series (2014) and the latest test series (2017), but the connection among all test series was strong.

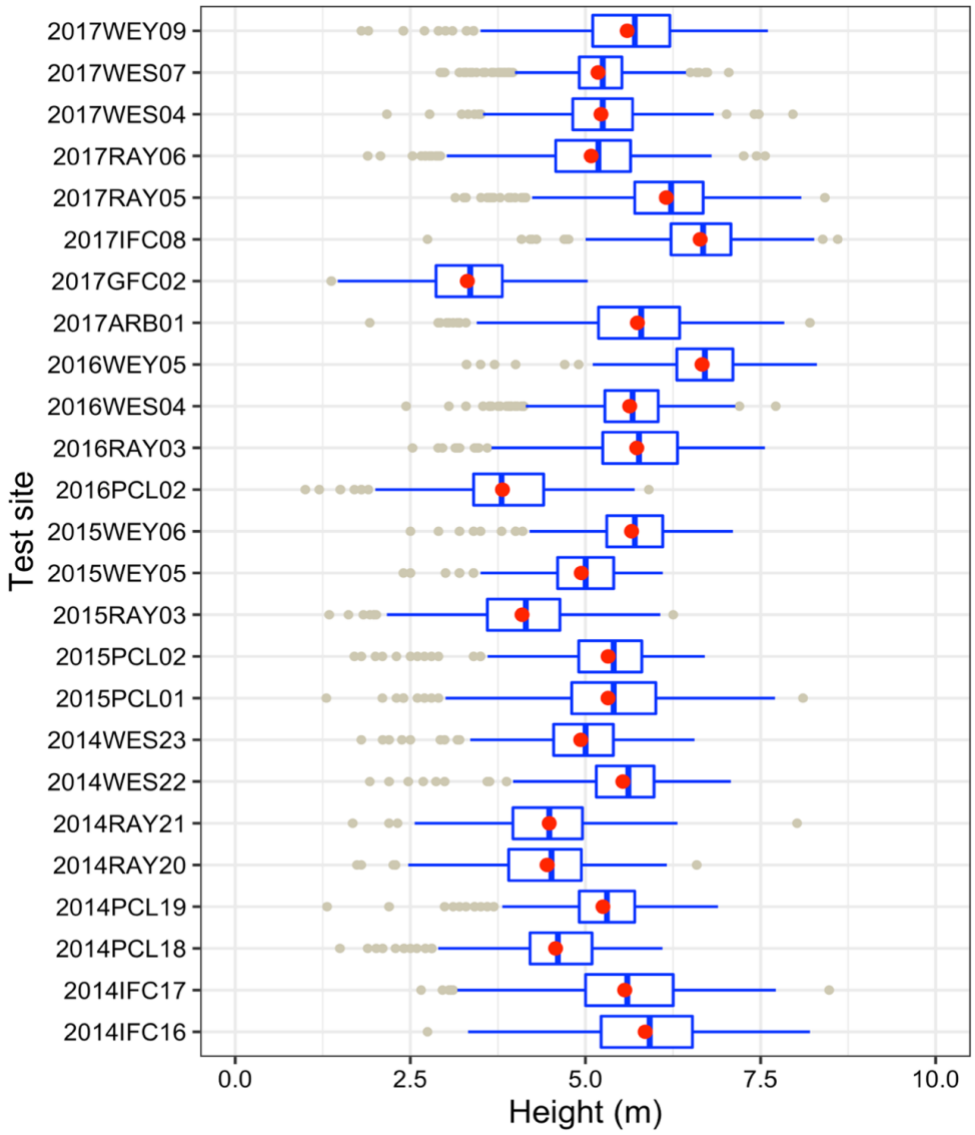

Figure 2.3 Height distribution at each test site. Red points are height means at each site while the heavy blue line inside the box is the median inside of the test site. The blue lines extending from two sides of the box present minimum and maximum height in that test site. The plot shows substantial site to site variation for height.

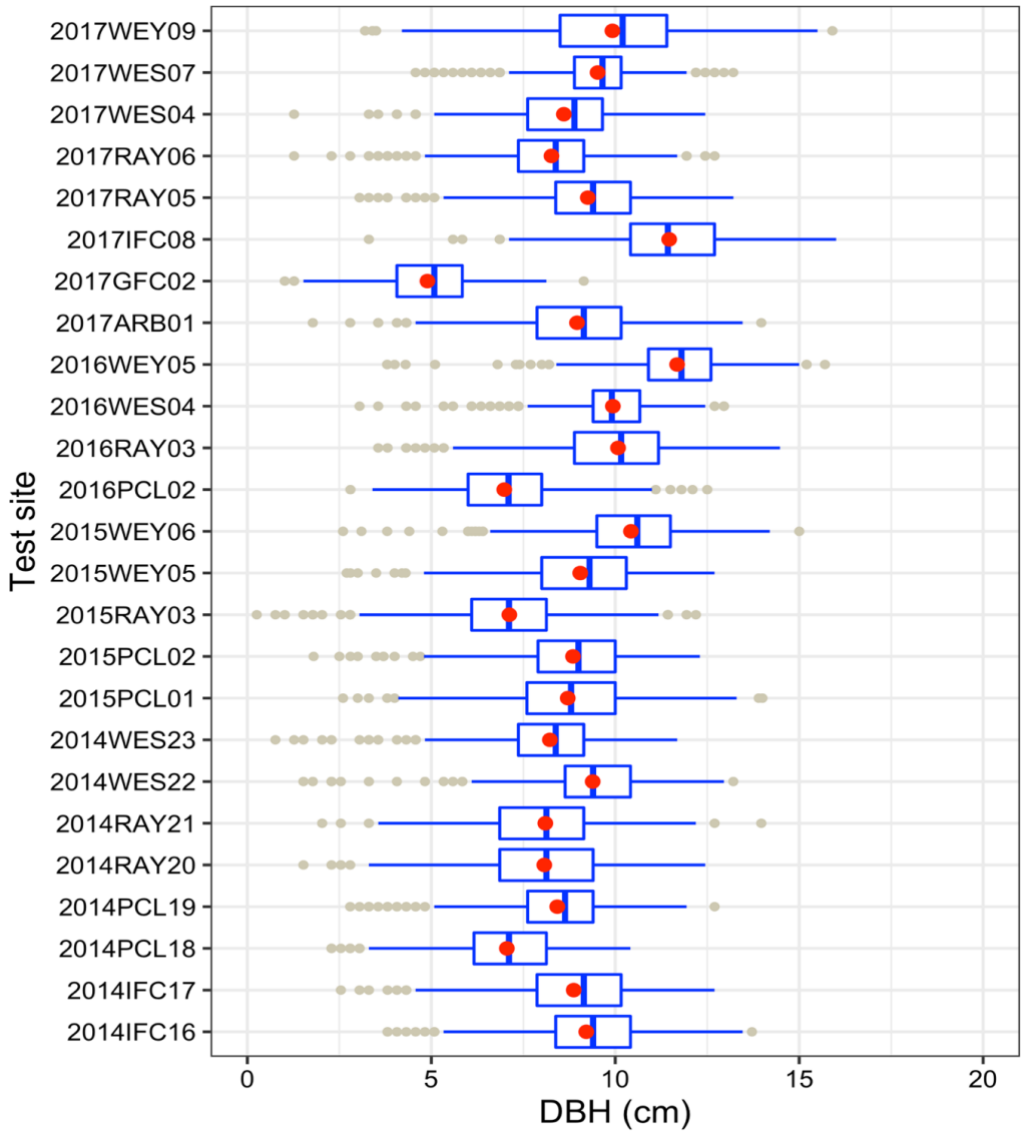

Figure 2.4 DBH distribution at each site. Red points are the site means while the heavy blue line in the box is the median. The blue lines extending from two side of the boxes present minimum and maximum values in that test site. The plots showed substantial site to site variation for DBH.

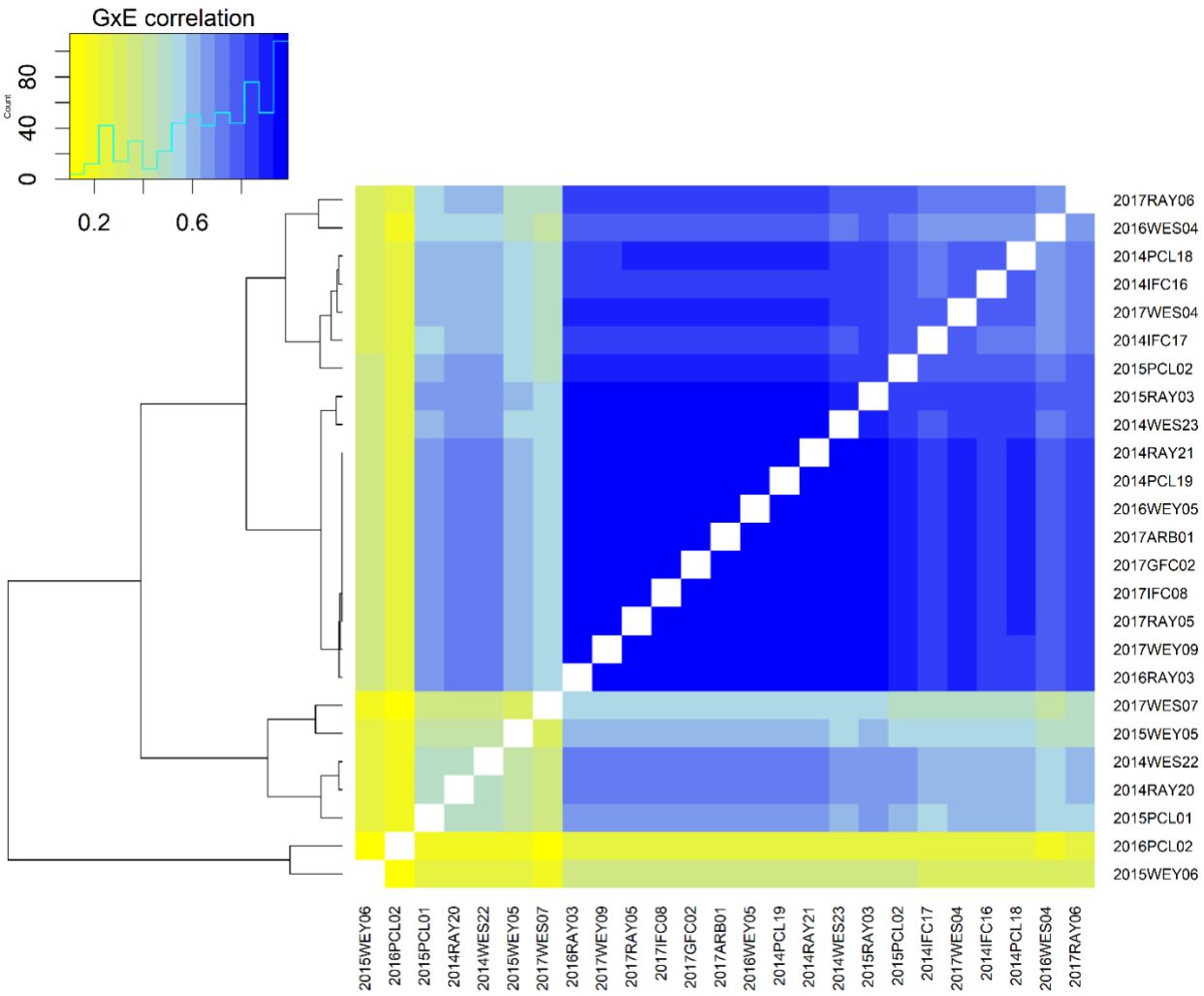

Figure 2.5 Dendrogram and heatmap of hierarchical clustering of additive genetic correlations (range between 0.10 and 0.99) of test sites from the XFA1 model for height. Blue color indicates high additive genetic correlations (low GxE), while yellow or pale blue indicate lower genetic correlations (higher GxE) among pairs of sites. Sites that are clustered together are more similar.

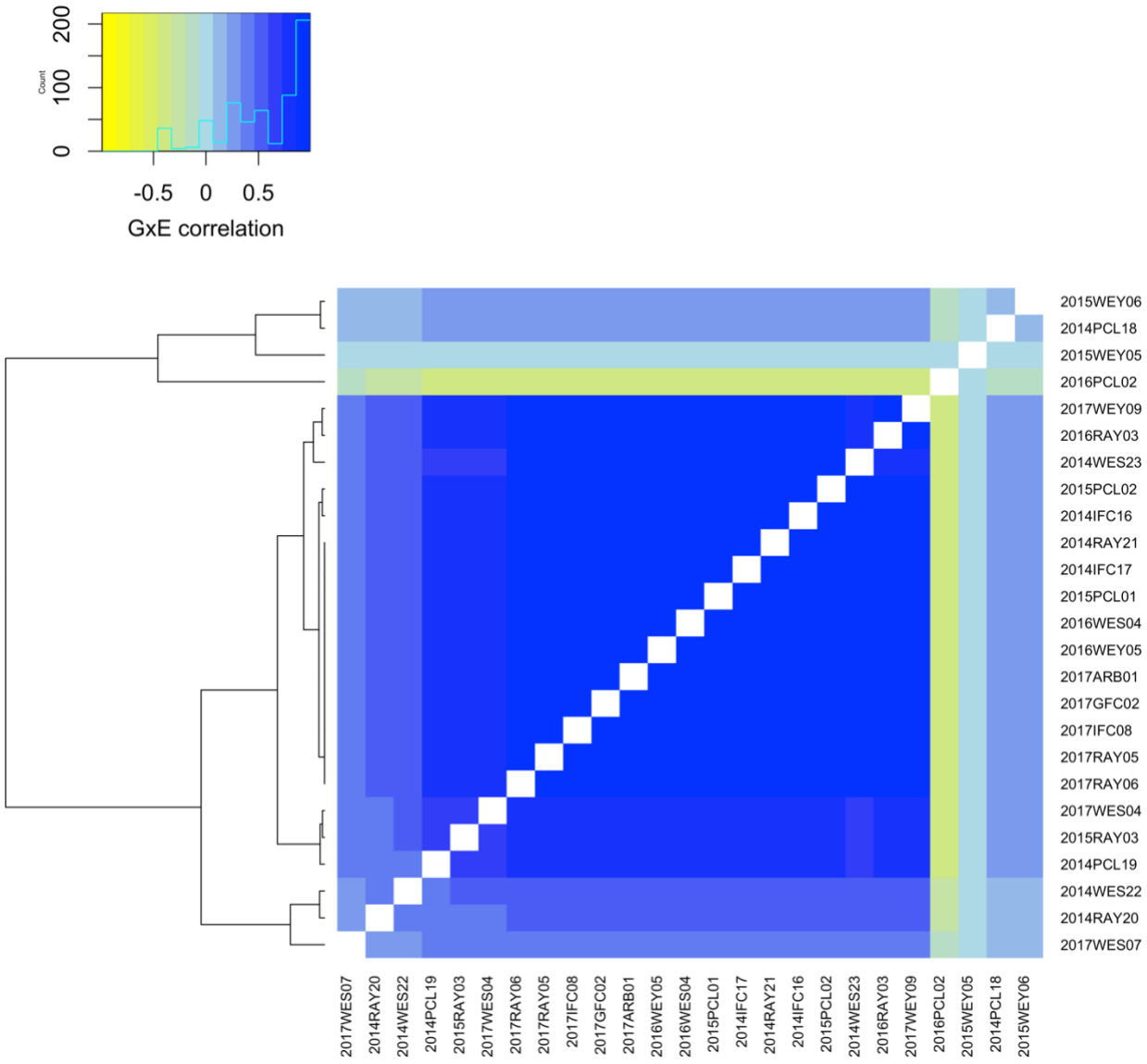

Figure 2.6 Dendrogram and heatmap of hierarchical clustering of additive genetic correlation (range between -0.44 and 0.99) of test sites from the XFA1 model for DBH. Blue color indicates high additive genetic correlations (low GxE) while yellow or pale blue indicate lower genetic correlations (higher GxE) among pairs of sites. Sites that are clustered together are more similar.

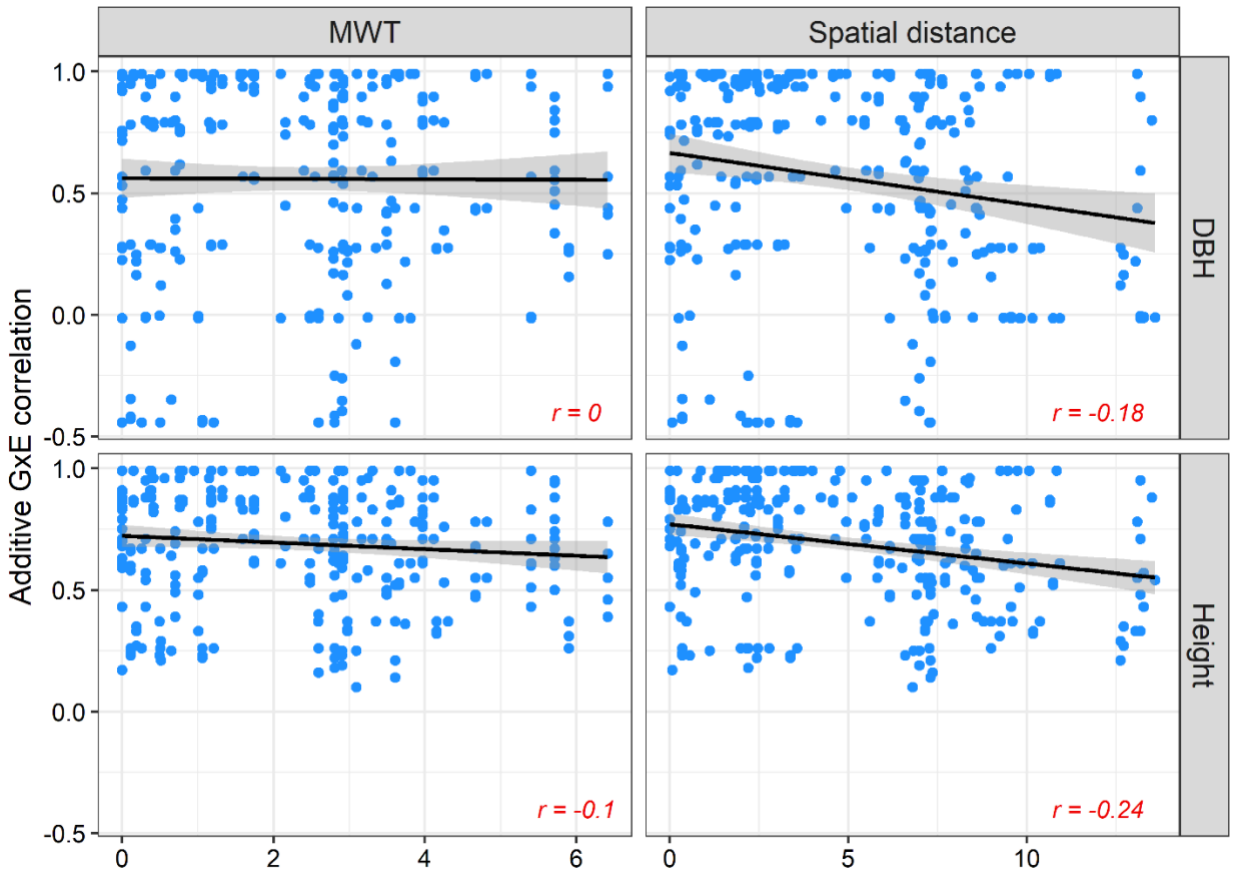

Figure 2.7 Scatterplots showing linear relationships between site-site additive genetic correlations for tree height and DBH (y-axis) with spatial distance (difference in latitude and longitude) and difference in minimum winter temperature (MWT) of site pairs. For both traits, larger differences in spatial distance are associated with lower additive genetic correlations.

Chapter 3. Genetic Parameter Estimates for Categorical Traits in *Pinus taeda* L. Fourth Cycle Coastal Breeding Population

3.1. Introduction

Fusiform rust disease in the southern pines of the United States is caused by *Cronartium quercuum f. sp. fusiforme*. The disease is an economically important problem of loblolly pine (*Pinus taeda*) and slash pine (*Pinus elliottii*) farming in the southern United States (McKeand et al., 1999; Kubisiak et al., 2005). Over millions of years, the pathogen has co-evolved with the host pine species in the region (Kubisiak et al., 2005). Fusiform rust disease has large implications on loblolly pine plantation management. In young pine plantations, the fungus causes annual losses of more than $100 million (Cubbage et al., 2000). The disease causes decreases in growth rate, reduced wood quality, and increased mortality rates in young pine plantations (Cubbage et al., 2000; Schmidt, 2003). Improvement for fusiform rust disease resistance is one of the major goals of loblolly pine tree breeding in the southern US (Nance et al., 1982; Schmidt, 2003).

Stem forking occurs when two or more stems of equal diameter divide the trunk of a tree, and it is another important threshold trait assessed in loblolly pine tree breeding programs (Xiong et al. 2013). Another stem defect observed in loblolly pine is ramicorn branching. A ramicorn branch has twice the diameter and a significantly steeper angle than other branches in the same whorl and can cause large knots in the stems and increase amounts of compression wood (Schermann et al., 1997; Xiong et al., 2014). Stem forking and ramicorn branches significantly affect wood quality and quantity, reducing the economic value of wood (Vargas-Hernandez et al.,

2003; Xiong et al., 2014). There is a common perception that environmental factors such as late-season frosts, wind damage, or silvicultural practices such as heavy thinning and wide spacing have a significant role in forking (Zobel & Kellison, 1978; Savill et al., 1999; Xiong et al., 2010).

Both fusiform rust disease incidence and stem forking or ramicorn branching are binary traits with a dichotomous phenotypic character (yes or no). For genetic analysis of binary traits, an underlying continuous distribution on a liability scale is assumed (Lynch & Walsh, 1998; Falconer & McKay 1995). When the underlying variable is above the threshold level, an individual tree is presumed to have one phenotype (e.g., forked or diseased), and vice versa if below the threshold level (e.g., not forked or not diseased) (Xiong et al., 2010; Isik et al., 2012). The underlying liability is assumed to be normally distributed.

Fusiform rust disease incidence in loblolly pine has been widely studied. However, information is limited on stem forking defects in the published literature as compared to fusiform rust disease incidence. A growing body of literature suggests fusiform rust disease incidence is under strong additive genetic control and classical recurrent selection will be effective to improve disease resistance in loblolly pine (McKeand et al., 1999; Isik et al., 2008; Spitzer et al., 2017; Shalizi et al., 2021). Moreover, families resistant to fusiform rust disease are the most stable across different environments, suggesting that GxE for fusiform rust incidence is not a concern in loblolly pine (McKeand et al., 2003, 2006; Isik et al., 2008).

Stem forking in loblolly pine is considered less heritable compared to fusiform rust disease resistance at the individual-tree level. Heritability of means for binary traits, such as stem forking

and fusiform rust disease incidence are more reliable and meaningful in a selection program (Isik et al., 2017). In a cloned loblolly pine experiment, moderate (0.67) clone mean repeatability for ramicorn branches was reported while the repeatability of clone means for forking defects was high (0.86) (Xiong et al., 2010). In another study in loblolly pine, narrow-sense individual tree heritability for stem forking was 0.15, while family mean heritability was high (0.92) (Cumbie et al., 2012).

The methods for assessing genetics of threshold traits are more complicated than those used in continuous traits. Generalized linear mixed model is one of the commonly used statistical methods for the analysis of threshold traits in breeding program (Gianola & Foulley, 1983). These models have been regularly used for the genetic analyses of fusiform rust disease incidence and stem forking in loblolly pine (Isik et al., 2012; Spitzer et al., 2017; Shalizi et al., 2021). In this study, we implemented such models to evaluate genetic variation and predict genetic merit of parents for stem forking defects and fusiform rust disease incidence in a large breeding population of loblolly pine.

The objectives of this study were to estimate genetic parameters and calculate individual-tree heritability for fusiform rust disease incidence and stem forking defects in the 4th-Cycle loblolly pine (*Pinus taeda L.*) Coastal breeding population.

3.2. Material and Methods

3.2.1 Genetic material

In this study, we analyzed 461 parents and their pedigrees belonging to three breeding populations. For each breeding population, the parents were mated using MateSel software to produce full-sib families (Kinghorn, 2011). In total, there were 376 full-sib families in the dataset. Full-sib families were represented by 60 progeny tested across at least two test series (years). At the time of assessment 17,775 trees were assessed for fusiform rust disease incidence.

3.2.2 Experimental design

The experimental design for all datasets analyzed was an alpha-cyclic incomplete row-column design (John & Williams, 1995; Williams et al., 2006). In this design, the rows and columns are regarded as incomplete blocks. Seedlings from each cross were assigned to specific rows and columns to optimize testing efficiency and improve direct comparisons among crosses. Seedling location in one rep is linked to seedling location in all reps throughout a testing series for a given year, both at the same site and at other sites. Families were typically tested across four years between 2014 and 2017. To provide connection and comparison of tests across years, common families were included in the testing strategy across multiple years. Test sites and test series were strongly connected with each other. The number of shared parents ranged between 38 and 164 among site pairs.

3.2.3 Data collection

Depending on the growth, progeny tests were assessed at ages four or five. Most tests were assessed at age four. The incidence of fusiform rust disease incidence was recorded as presence or absence of disease galls. Similarly, stem forking and ramicorn branching were recorded as yes or no. These two traits are the major defects for stems and timber quality. Stem forking and ramicorn branching were later combined as one trait for further analysis. If either or both were present, a tree was considered forked since the two traits are genetically related (Xiong et al., 2010). R statistical software was used for data cleaning, visualization, descriptive and summary statistics (R development Core Team, 2021).

3.2.4 Statistical Analysis

Fusiform rust disease incidence and stem forking incidence were analyzed using the following generalized mixed model to estimate variance components and genetic merits of 294 unique parents and 376 crosses.

$$\eta = \log\left[\frac{p}{1-p}\right] = \mu + \boldsymbol{X}s + \boldsymbol{Z}_1 b + \boldsymbol{Z}_2 q + \boldsymbol{Z}_3 c + \boldsymbol{Z}_4 m + \boldsymbol{Z}_6 f + \boldsymbol{Z}_7 i + \boldsymbol{Z}_8 j + e \qquad (3.1)$$

where, η is the link function of the vector of response variable (incidence / no incidence); p is the probability of the trait incidence; $\log[p/(1-p)]$ is the log of odds; μ is the conditional mean; $\boldsymbol{X}$ and $\boldsymbol{Z}$ are incidence matrices for the fixed and random effects, respectively; s is the vector of site fixed effect ($s = 25$); b is the vector for random replicate effect nested within site with $b \sim MVN(0, \boldsymbol{I}\sigma_b^2)$; q is the vector of random row effect nested within rep with $q \sim MVN(0, \boldsymbol{I}\sigma_q^2)$; c is the vector of random column effect nested within rep with $c \sim MVN(0, \boldsymbol{I}\sigma_c^2)$; m is the vector of

random GCA (general combining ability) female and male effects with $m \sim MVN(0, \boldsymbol{A}\sigma_m^2)$; f is the vector of random family effects with $f \sim MVN(0, \boldsymbol{I}\sigma_f^2)$; i is the vector of random GCA female and male by site effect with $i \sim MVN(0, \boldsymbol{A}\sigma_i^2)$; j is the vector of random family by site effect with $j \sim MVN(0, \boldsymbol{I}\sigma_j^2)$; and e is the vector of random errors with $e \sim MVN(0, \boldsymbol{I}\sigma_e^2)$ where, $\boldsymbol{I}$ is an identity matrix of its proper dimensions, $\boldsymbol{A}$ is the numerator relationship matrix calculated from pedigree. The pedigree was comprised of grandparents, parents, and 294 females and males.

Individual-tree narrow-sense (h_i^2) and broad-sense (H_i^2) heritabilities were estimated for the two traits using the derivations of the equations according to Isik et al. (Isik et al., 2017).

$$h_i^2 = \frac{4\sigma_m^2}{2\sigma_m^2+\sigma_f^2+2\sigma_i^2+\sigma_j^2+\sigma_e^2} \quad (3.2)$$

$$H_i^2 = \frac{4\sigma_m^2+4\sigma_f^2}{2\sigma_m^2+\sigma_f^2+2\sigma_i^2+\sigma_j^2+\sigma_e^2} \quad (3.3)$$

where, σ_m^2 is the GCA female and male variance; σ_f^2 is the family variance; σ_i^2 is the GCA female and male by site variance; σ_j^2 is the family by site variance; and σ_e^2 is the residual variance. The additive and dominance genetic variances were obtained as $4\sigma_m^2$ and $4\sigma_f^2$, respectively. The variance of a standard binomial distribution ($\pi^2/3 = 3.29$) was used as the error variance (Gilmour et al., 1985). The standard error of heritabilities were calculated using delta method (Lynch & Walsh, 1998).

The solutions of mixed model equations for parents and for full-sib families were obtained. These estimates are the best linear unbiased estimates (BLUP) on the logit scale. They were converted to the probability scale on the mean incidence using the exponential function in ASReml Predict statement (Gilmour et al., 2015). The predicted breeding values on the probability scale of parents and full-sib families were visualized with bar charts using the *ggplot2* package of R (Wickham, 2010).

3.3. Results

3.3.1. Summary statistics

Summary statistics for test sites for fusiform rust disease incidence and stem forking are presented in **Table 3.1.** The incidence of the two traits averaged 22% across all sites. The incidence of fusiform rust disease ranged between 3% and 49%, while stem forking ranged between 4% and 48% among the test sites. There was large variation for observed fusiform rust disease incidence in the population. On the full-sib family level, the incidence of fusiform rust disease incidence ranged between 0 and 80%, while stem forking ranged between 0 and 46% among families (**Figure 3.1**).

3.3.2. Variance components and heritability

Genetic parameter estimates for fusiform rust disease and stem forking incidence calculated from GCA models are presented in **Table 3.2**. Additive genetic variance was significant (0.97±0.154) and was 3.7 times greater than the dominance genetic variance for fusiform rust disease incidence. Additive genetic variance for stem forking was 0.13±0.033 and explained a small proportion (4%) of the phenotypic variance. Dominance genetic variance for fusiform rust disease incidence was considerable (0.26 ±0.110), but dominance genetic variance for stem forking was essentially zero. The GCA-by-site and SCA-by-site variances were small in both traits, explaining less than 1% of the phenotypic variance.

Narrow-sense and broad-sense heritability estimates were considerably higher for fusiform rust disease incidence compared to heritabilities for stem forking. Narrow-sense and broad-sense heritability estimates were 0.25±0.035 and 0.31±0.033 for fusiform rust disease incidence,

respectively (**Table 3.2**). The narrow-sense and broad-sense heritability for stem forking were 0.04±0.009. The additive GxE correlation was high (0.98) for the two traits indicating that parents rank consistently across different environments for these two traits.

Fusiform rust disease incidence predicted breeding values of the 294 parents (females and males), and genetic values of the 376 full-sib families are presented in **Figure 3.2** and **Figure 3.3**. Large variation among both parents and full-sib families indicated substantial genetic variation for fusiform rust disease resistance. The breeding values of 294 parents ranged between 0.07 to 0.91, whereas the genetic values of the 376 families ranged from 0.14 to 0.86 (**Figure 3.2** and **3.3**). For stem forking, the range of predictions (0.09 to 0.33) for parental breeding values was narrower, compared to fusiform rust disease incidence (**Figure 3.4**). Similarly, the probability of forking for full-sib families had a narrow range (0.12 to 0.25) compared to fusiform rust disease incidence (**Figure 3.5**).

3.4. Discussion

In this study, additive genetic variance explained a significant amount of the total genetic variance for fusiform rust disease incidence. We observed considerable genetic variation between parents (probability of disease 0.07 to 0.91) and full-sib families (probability of disease 0.14 to 0.86) for fusiform rust disease incidence as indicated by a wide range of predictions for both parents and full-sib families. High heritability estimates of fusiform rust disease incidence suggests that the trait is substantially controlled by genetics. The heritability estimates of fusiform rust disease incidence from this study were comparable to the heritability estimates for a growing number of published studies in loblolly pine (Isik & Li, 2003; Isik et al., 2005; Cumbie et al., 2012; Spitzer et al., 2017; Shalizi, 2020).

Fusiform rust disease incidence follows classical Mendelian segregation in QTL mapping studies when host parents are challenged with single spore isolates (Wilcox et al., 1996; Lauer & Isik, 2021). However, the inheritance of the trait is more complex at the population level, as we observed in this study. There are likely many more QTLs in different families segregating among. These QTL may interact with different diverse spore isolates in the natural environment (Kubisiak et al., 2005; Isik et al., 2012). Plus, environmental factors such as temperature and humidity can alter disease occurrence at a given environment. These multiple factors influence the expression of the trait causing it to behave like a polygenic trait at the population level that responds well to classical truncation selection (Isik et al., 2008).

Stem forking seemed to be largely controlled by environmental factors as indicated by small additive genetic variance and low heritability estimates. A previously published study

reported low individual-tree heritability estimates for stem forking in loblolly pine (Xiong et al., 2010). They concluded that stem forking is under weak genetic control on an individual-tree level. Some studies on loblolly pine indicated that stem forking is largely controlled by genetics at family level (Xiong et al., 2010; 2013). There is a common perception that environmental factors such as late-season frosts, tip moth damage, wind damage, silvicultural practices such as heavy thinning, and wide spacing can have a significant effect on stem forking occurrence (Zobel & Kellison, 1978; Savill et al., 1999; Xiong et al., 2010).

There was no GxE interaction for fusiform rust disease incidence and stem forking incidence in the 4th-Cycle Coastal population. In a study in the southern USA, families that were resistant to fusiform rust disease were also the most stable across different environments (McKeand et al., 1999, 2003). The results confirm findings of previously published research indicating stable performance of genotypes of the two traits across multiple environments in the southern US (Isik et al., 2005; Cumbie et al., 2012; Spitzer et al., 2017).

Stem forking defects and fusiform rust disease incidence should be included in selection decisions moving forward. However, different selection strategies might be employed for fusiform rust disease incidence and stem forking. Fusiform rust disease incidence could be included in a multi-trait index selection with a weight to simultaneously improve disease resistance, growth traits, and stem straightness. For stem forking, a threshold selection model could be implemented, where a proportion of full-sib families and parents above a threshold value, such as parents with probability of 0.70 for forking could be removed before a selection index is constructed.

3.5. Conclusions

The analysis of the 4th-Cycle Coastal progeny test data indicated large additive genetic variation for fusiform rust disease incidence. We detected a large number of parents and full-sib families resistant to fusiform rust disease in the Coastal population. The GxE for fusiform rust was essentially zero, confirming the published results on the subject. Stem forking was weakly controlled by additive genetic effects. The trait is likely controlled more by environmental factors, such as frost/ice damage or biotic factors. Still, a moderate truncation selection strategy (culling high forking crosses/parents) at the family level will be effective.

3.6. References

Cubbage, F. W., Pye, J. M., Holmes, T. P., & Wagner, J. E. (2000). An Economic Evaluation of Fusiform Rust Protection Research. *Southern Journal of Applied Forestry*, *24*(2), 77–85. https://doi.org/10.1093/sjaf/24.2.77

Cumbie, W. P., Isik, F., & McKeand, S. E. (2012). Genetic Improvement of Sawtimber Potential in Loblolly Pine. *Forest Science*, *58*(2), 168–177. https://doi.org/10.5849/forsci.09-060

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.).

Gianola, D., & Foulley, J. L. (1983). *Sire evaluation for ordered categorical data with a threshold model*. 23.

Gilmour, A. R., Anderson, R. D., & Rae, A. L. (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika*, *72*(3), 593–599. https://doi.org/10.1093/biomet/72.3.593

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., & Thompson, R. (2015). ASReml user guide release 4.1 functional specification. *Hemel Hempstead: VSN International Ltd*.

Isik, F., Amerson, H. V., Whetten, R. W., Garcia, S. A., Li, B., & McKeand, S. E. (2008). Resistance of Pinus taeda families under artificial inoculations with diverse fusiform rust pathogen populations and comparison with field trials. *Canadian Journal of Forest Research*, *38*(10), 2687–2696. https://doi.org/10.1139/X08-111

Isik, F., Amerson, H. V., Whetten, R. W., Garcia, S. A., & McKeand, S. E. (2012). Interactions of Fr genes and mixed-pathogen inocula in the loblolly pine-fusiform rust pathosystem. *Tree Genetics & Genomes*, *8*(1), 15–25. https://doi.org/10.1007/s11295-011-0416-0

Isik, F., Boos, D. D., & Li, B. (2005). The distribution of genetic parameter estimates and confidence intervals from small disconnected diallels. *Theoretical and Applied Genetics*, *110*(7), 1236–1243. https://doi.org/10.1007/s00122-005-1957-0

Isik, F., Holland, J., & Maltecca, C. (2017). *Genetic Data Analysis for Plant and Animal Breeding*. Springer International Publishing. https://doi.org/10.1007/978-3-319-55177-7

Isik, F., & Li, B. (2003). *Variation in estimation of genetic parameters from small disconnected diallel mating*. *27*, 20.

John, J. A., & Williams, E. R. (1995). *Cyclic and computer generated designs. Monographs on Statistics and Applied Probability 38.* Chapman & Hall.

Kinghorn, B. P. (2011). An algorithm for efficient constrained mate selection. *Genetics Selection Evolution*, *43*(1), 4. https://doi.org/10.1186/1297-9686-43-4

Kubisiak, T. L., Amerson, H. V., & Nelson, C. D. (2005). Genetic Interaction of the Fusiform Rust Fungus with Resistance Gene *Fr* 1 in Loblolly Pine. *Phytopathology*®, *95*(4), 376–380. https://doi.org/10.1094/PHYTO-95-0376

Lauer, E., & Isik, F. (2021). Major QTL confer race-nonspecific resistance in the co-evolved Cronartium quercuum f. Sp. Fusiforme—Pinus taeda pathosystem. *Heredity*.

Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates Incorporated. https://www.journals.uchicago.edu/doi/10.1086/393101

McKeand, S. E., Abt, R. C., Allen, H. L., Li, B., & Catts, G. P. (2006). What Are the Best Loblolly Pine Genotypes Worth to Landowners? *Journal of Forestry*, *104*(7), 352–358.

McKeand, S. E., Amerson, H. V., Li, B., & Mullin, T. J. (2003). Families of loblolly pine that are the most stable for resistance to fusiform rust are the least predictable. *Canadian Journal of Forest Research*, *33*(7), 1335–1339. https://doi.org/10.1139/x03-050

McKeand, S. E., Li, B., & Amerson, H. V. (1999). Genetic variation in fusiform rust resistance in loblolly pine across a wide geographic range. *Silvae Genetica*, *48*(5), 255–259.

Nance, W. L., Froelich, R. D., Dell, T. R., & Shoulders, E. (1982). *A growth and yield model or unthinned slash pine plantations infected with fusiform rust*. 2nd Biennial South. Silvic. Res. Conf., U.S. Dep. Agric. For. Serv. Atlanta, GA.

R development Core Team. (2021). *R: A language and environment for statistical computing* (4.1.1) [Computer software]. R Foundation for Statistical Computing.

Savill, P. S., Spencer, R., Roberts, J. E., & Hubert, J. D. (1999). *Sixth Year Results from Four Ash (Fraxinus excelsior) Breeding Seedling Orchards*. 9.

Schermann, N., Adams, W. T., & Aitken, S. N. (1997). *Genetic Parameters of Stem Form Traits in a 9-Year-Old Coastal Douglas-fir Progeny Test in Washington 1)*. 5.

Schmidt, R. A. (2003). Fusiform Rust of Southern Pines: A Major Success for Forest Disease Management. *Phytopathology*®, *93*(8), 1048–1051. https://doi.org/10.1094/PHYTO.2003.93.8.1048

Shalizi, M. N. (2020). *Genetic Parameter Estimates and Prediction of Genetic Merit in Clonal Populations of Pinus taeda L. using Expected and Realized Genomic Relationships* [NCSU].https://repository.lib.ncsu.edu/bitstream/handle/1840.20/38351/etd.pdf?sequence=1&isAllowed=y

Shalizi, M. N., Cumbie, W. P., & Isik, F. (2021). Genomic prediction for fusiform rust disease incidence in a large cloned population of Pinus taeda. *G3 Genes|Genomes|Genetics*, *11*(9). https://doi.org/10.1093/g3journal/jkab235

Spitzer, J., Isik, F., Whetten, R. W., Farjat, A. E., & McKeand, S. E. (2017). Correspondence of Loblolly Pine Response for Fusiform Rust Disease from Local and Wide-Ranging Tests

in the Southern United States. *Forest Science*, *63*(5), 496–503. https://doi.org/10.5849/FS-2016-093R1

Vargas-Hernandez, J. J., Adams, W. T., & Joyce, D. G. (2003). *Quantitative Genetic Structure of Stem Form and Branching traits in Douglas-fir Seedlings and Implications for Early Selection*. 9.

Wickham, H. (2010). *ggplot2: Elegant Graphics for Data Analysis* (1st ed. 2009. Corr. 3rd printing 2010 edition). Springer.

Wilcox, P. L., Amerson, H. V., Kuhlman, E. G., Liu, B. H., O'Malley, D. M., & Sederoff, R. R. (1996). Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(9), 3859–3864.

Williams, E. R., John, J. A., & Whitaker, D. (2006). Construction of Resolvable Spatial Row–Column Designs. *Biometrics*, *62*(1), 103–108. https://doi.org/10.1111/j.1541-0420.2005.00393.x

Xiong, J. S., Isik, F., McKeand, S. E., & Whetten, R. W. (2010). *Genetic Variation of Stem Forking in Loblolly Pine*. 8.

Xiong, J. S., McKeand, S. E., Whetten, R. W., & Isik, F. T. (2013). Genetics of stem forking and ramicorn branches in a cloned loblolly pine family. *Forest Science*, *60*(2), 360–366.

Xiong, J. S., McKeand, S. E., Whetten, R. W., & Isik, F. T. (2014). Genetics of Stem Forking and Ramicorn Branches in a Cloned Loblolly Pine Family. *Forest Science*, *60*(2), 360–366. https://doi.org/10.5849/forsci.12-018

Zobel, B. J., & Kellison, R. C. (1978). The rate of growth syndrome. *Silvae Genetica*, *27*(3/4), 123–124.

Table 3.1 Summary statistics, number of observations (N), mean (±standard error) for fusiform rust disease incidence and stem forking incidence for the 4th-Cycle Coastal test sites.

Site	Location (County, State)	N	Rust ±SE	Forking ±SE
2014IFC16	COLQUITT, GA	421	0.49 ±0.024	0.18 ±0.019
2014IFC17	COLQUITT, GA	424	0.49 ±0.024	0.16 ±0.018
2014PCL18	BULLOCH, GA	423	0.42 ±0.024	0.15 ±0.018
2014PCL19	BULLOCH, GA	420	0.42 ±0.024	0.17 ±0.019
2014RAY20	NASSAU, FL	411	0.38 ±0.024	0.04 ±0.010
2014RAY21	NASSAU, FL	427	0.37 ±0.023	0.12 ±0.015
2014WES22	GREENE, AL	441	0.24 ±0.020	0.41 ±0.023
2014WES23	GREENE, AL	464	0.27 ±0.020	0.38 ±0.023
2015PCL01	EFFINGHAM, GA	611	0.22 ±0.017	0.10 ±0.012
2015PCL02	BULLOCH, GA	625	0.30 ±0.018	0.07 ±0.010
2015RAY03	LONG, GA	1196	0.25 ±0.012	0.09 ±0.008
2015WEY05	CRAVEN, NC	484	0.29 ±0.020	0.29 ±0.021
2015WEY06	JONES, NC	535	0.31 ±0.020	0.31 ±0.020
2016PCL02	EFFINGHAM, GA	817	0.10 ±0.010	0.06 ±0.008
2016RAY03	NASSAU, FL	812	0.19 ±0.014	0.36 ±0.017
2016WES04	PICKENS, AL	1643	0.08 ±0.007	0.32 ±0.012
2016WEY05	CRAVEN, NC	835	0.43 ±0.017	0.19 ±0.013
2017ARB01	SANTA ROSA, FL	795	0.09 ±0.010	0.16 ±0.013
2017GFC02	DOOLY, GA	560	0.09 ±0.012	0.48 ±0.021
2017IFC08	COLQUITT, GA	673	0.21 ±0.016	0.32 ±0.018
2017RAY05	ATKINSON, GA	796	0.23 ±0.015	0.15 ±0.012
2017RAY06	NASSAU, FL	797	0.17 ±0.013	0.14 ±0.012
2017WES04	GREENE, AL	786	0.16 ±0.013	0.34 ±0.017
2017WES07	PICKENS, AL	1596	0.03 ±0.005	0.32 ±0.012
2017WEY09	BULLOCH, GA	807	0.22 ±0.015	0.08 ±0.009

Table 3.2 Additive variance (σ_a^2), dominance variance (σ_d^2), ratio between two (σ_a^2/σ_d^2), GCA by site variance (σ_i^2), SCA by site variance (σ_j^2), individual narrow-sense heritability (h^2), individual broad-sense heritability (H^2) for fusiform rust incidence.

Estimates	**Rust ±SE**	**Forking ±SE**
Additive variance (σ_a^2)	0.97 ±0.154	0.13 ±0.033
Dominance variance (σ_d^2)	0.26 ±0.110	0.00 ±0.000
σ_a^2/σ_d^2	3.68	>100
GCA by site variance (σ_i^2)	0.01 ±0.017	<0.01 ±0.001
SCA by site variance (σ_j^2)	0.03 ±0.044	0.04 ±0.037
Narrow-sense heritability (h_i^2)	0.25 ±0.035	0.04 ±0.009
Broad-sense heritability (H_i^2)	0.31 ±0.033	0.04 ±0.009
Additive genetic correlation (r_A)	0.98 ±0.017	0.98 ±0.099

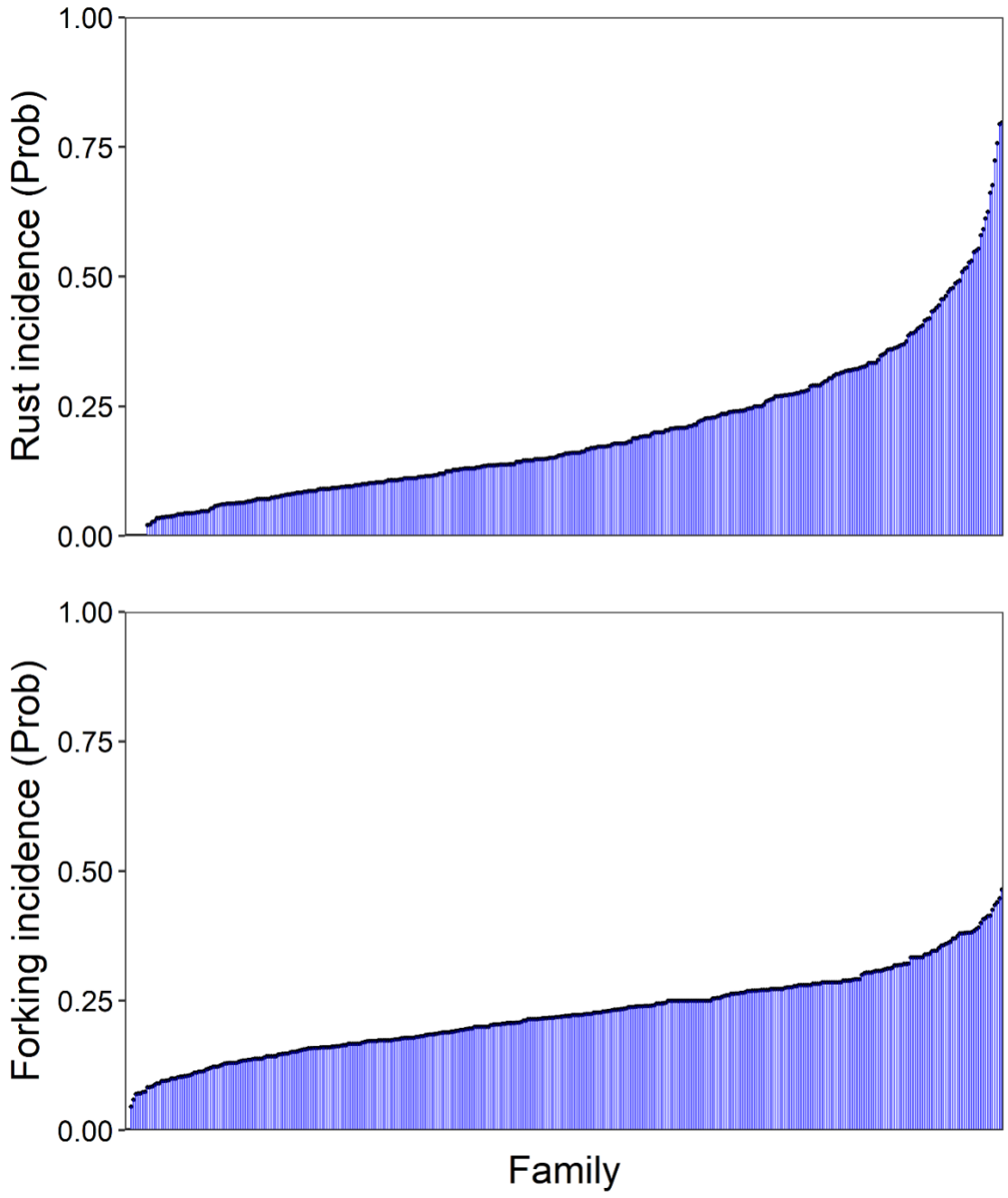

Figure 3.1 Mean incidence of observed fusiform rust disease and stem forking for the 376 full-sib families of loblolly pine in the 4th-Cycle Coastal population.

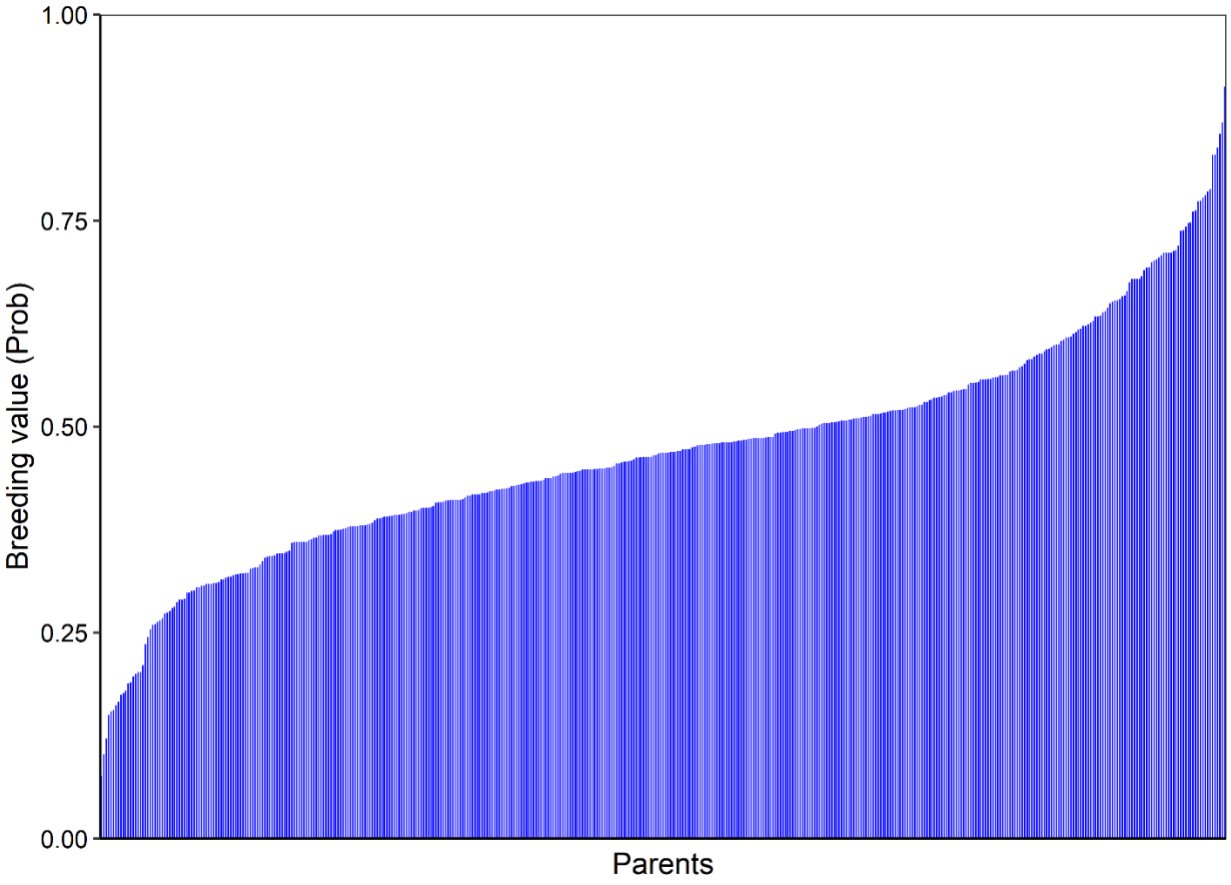

Figure 3.2 Breeding values on probability scale for 294 parents (female or male) evaluated in the 4th-Cycle Coastal population for fusiform rust disease incidence. Lower predicted breeding values suggest high resistance to fusiform rust disease.

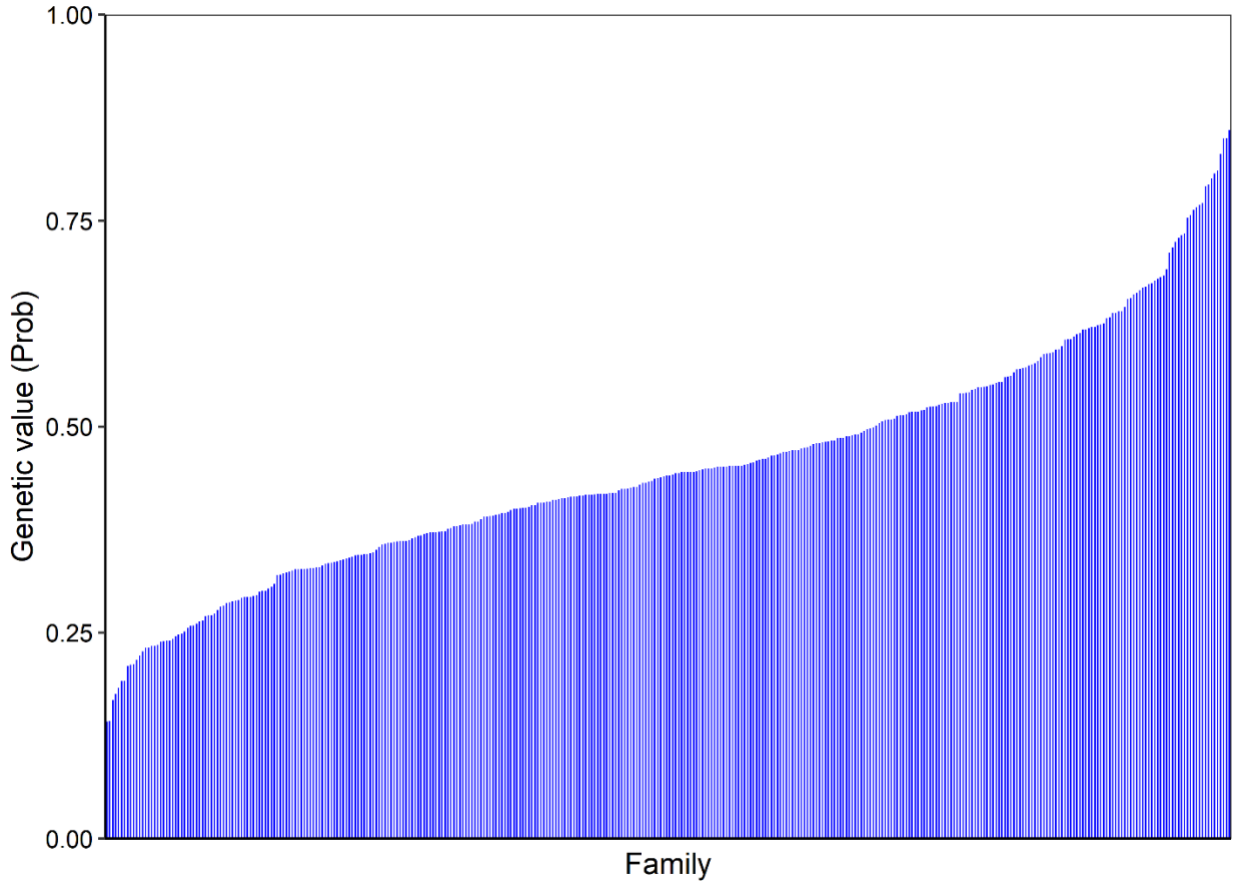

Figure 3.3 Genetic values on probability scale for 376 full-sib families evaluated in the 4th-Cycle Coastal population for fusiform rust disease incidence. Lower predicted breeding values suggest high resistance to fusiform rust disease.

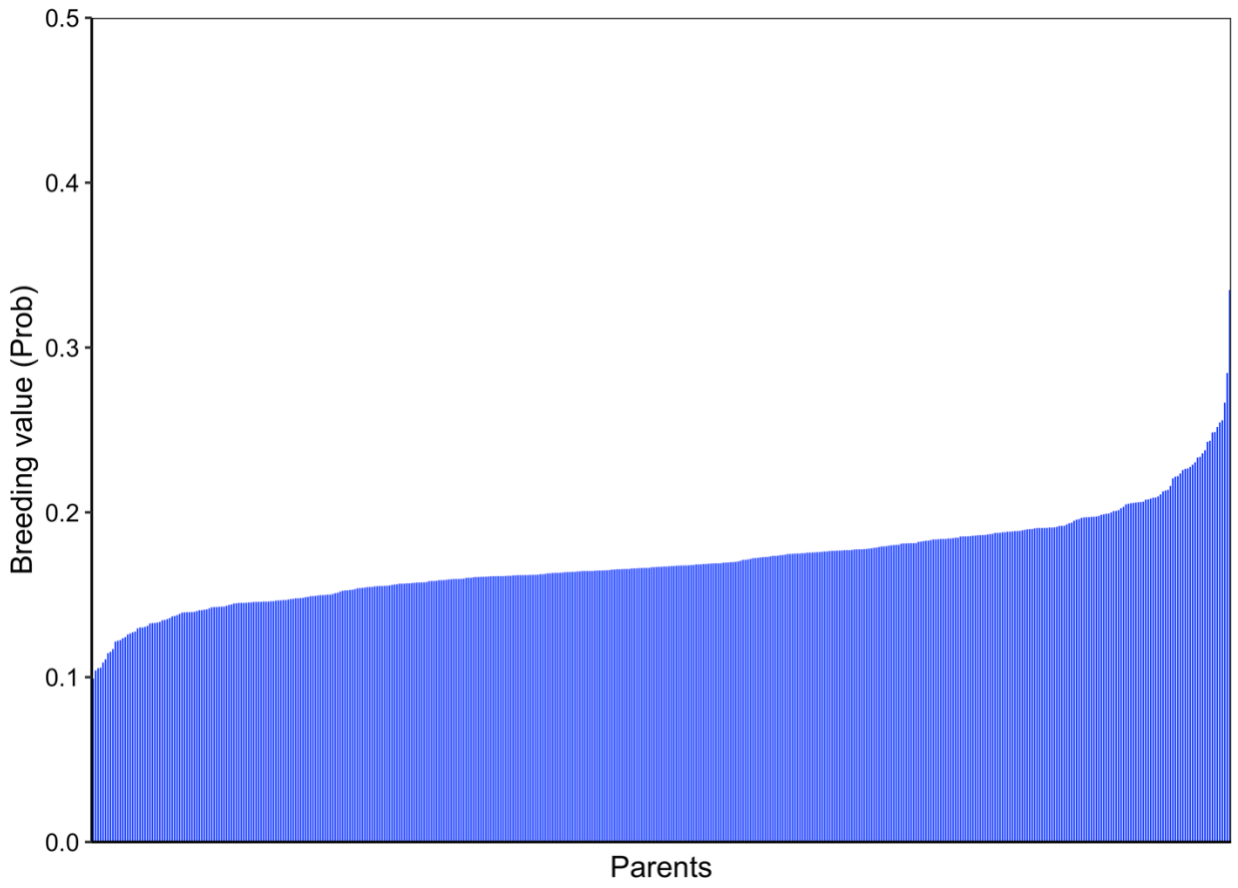

Figure 3.4 Breeding values for stem forking on probability scale for 294 parents (female or male) evaluated in the 4th-Cycle Coastal population. The range of predictions was 0.09 to 0.33 on the probability scale. Lower predicted breeding values suggest low probability of stem forking.

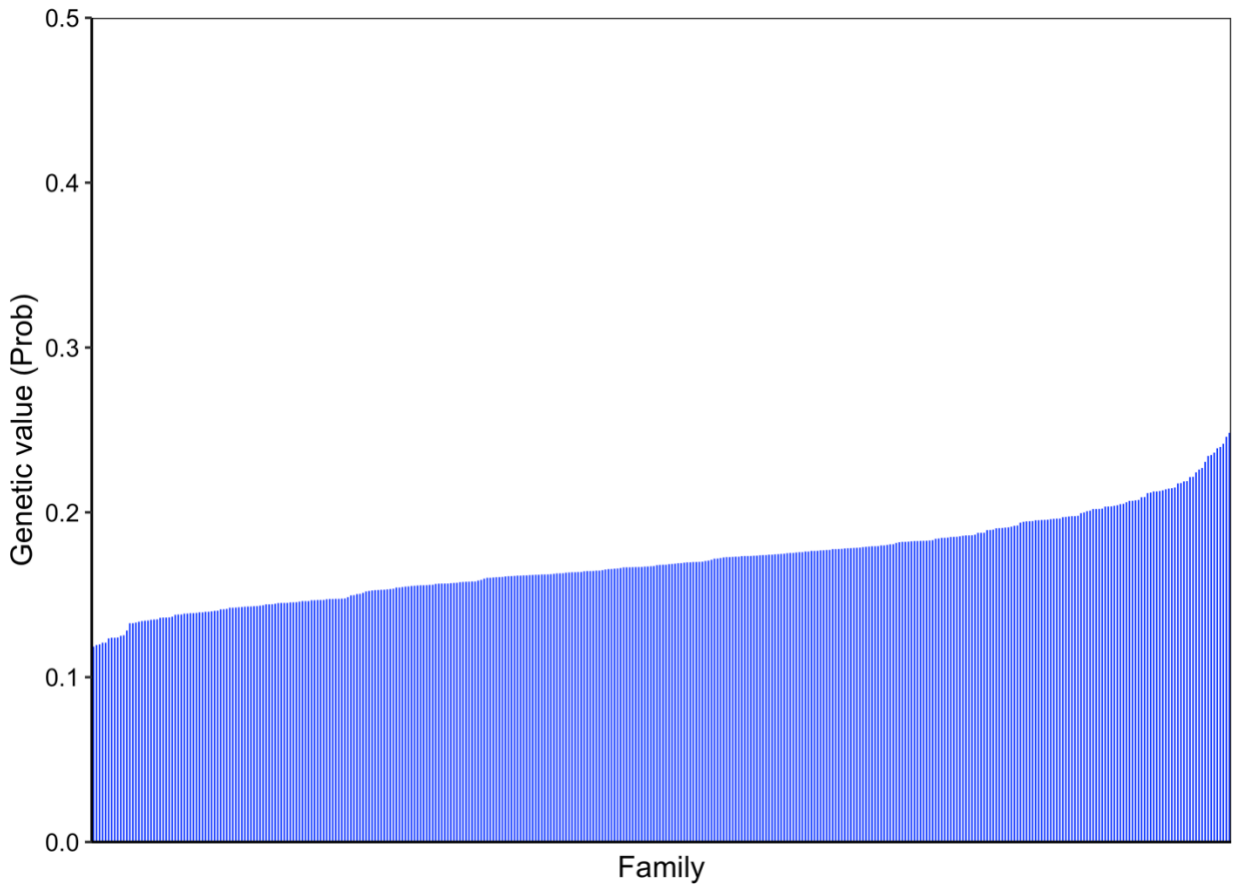

Figure 3.5 Genetic values on probability scale for 376 full-sib families evaluated in the 4th-Cycle Coastal population for stem forking. The range of predictions was 0.12 to 0.25 on the probability scale. Lower predicted breeding values suggest low probability of stem forking.

APPENDICES

Appendix A. Additional summary statistics for traits

Table A.2.1Summary statistics for height (m). Supplementary table for Figure 2.3.

Test	N	Mean	Std Dev	Std Error	Minimum	Maximum
2014IFC16	421	5.85	0.99	0.049	2.74	8.20
2014IFC17	424	5.56	0.98	0.048	2.65	8.48
2014PCL18	423	4.57	0.80	0.039	1.49	6.10
2014PCL19	420	5.25	0.67	0.033	1.31	6.89
2014RAY20	411	4.45	0.80	0.039	1.74	6.59
2014RAY21	427	4.48	0.78	0.038	1.68	8.02
2014WES22	441	5.53	0.71	0.034	1.92	7.08
2014WES23	464	4.93	0.69	0.032	1.80	6.56
2015PCL01	611	5.32	0.95	0.038	1.30	8.11
2015PCL02	625	5.32	0.73	0.029	1.70	6.70
2015RAY03	1196	4.10	0.77	0.022	1.34	6.25
2015WEY05	484	4.94	0.62	0.028	2.40	6.10
2015WEY06	535	5.66	0.66	0.029	2.50	7.10
2016PCL02	817	3.82	0.73	0.026	1.00	5.90
2016RAY03	812	5.73	0.80	0.028	2.53	7.56
2016WES04	1643	5.63	0.58	0.014	2.44	7.72
2016WEY05	835	6.66	0.65	0.022	3.30	8.31
2017ARB01	795	5.74	0.90	0.032	1.92	8.20
2017GFC02	560	3.31	0.67	0.028	1.37	5.03
2017IFC08	673	6.64	0.67	0.026	2.74	8.60
2017RAY05	796	6.15	0.78	0.028	3.14	8.42
2017RAY06	797	5.08	0.82	0.029	1.89	7.56
2017WES04	786	5.22	0.65	0.023	2.17	7.96
2017WES07	1596	5.18	0.52	0.013	2.93	7.05
2017WEY09	807	5.60	0.88	0.031	1.80	7.60

Table A.2.2 Summary statistics for diameter at breast height (cm). Supplementary table for Figure 2.4.

Test	N	Mean	Std Dev	Std Error	Minimum	Maximum
2014IFC16	421	9.21	1.75	0.085	3.81	13.72
2014IFC17	424	8.87	1.72	0.084	2.54	12.70
2014PCL18	423	7.05	1.60	0.078	2.29	10.41
2014PCL19	420	8.42	1.59	0.078	2.79	12.70
2014RAY20	411	8.07	1.87	0.092	1.52	12.45
2014RAY21	427	8.10	1.84	0.089	2.03	13.97
2014WES22	441	9.39	1.64	0.078	1.52	13.21
2014WES23	464	8.22	1.68	0.078	0.76	11.68
2015PCL01	611	8.71	1.81	0.073	2.60	14.00
2015PCL02	625	8.85	1.60	0.064	1.80	12.30
2015RAY03	1196	7.12	1.69	0.049	0.25	12.19
2015WEY05	484	9.05	1.72	0.078	2.70	12.70
2015WEY06	535	10.43	1.67	0.072	2.60	15.00
2016PCL02	817	6.98	1.56	0.055	2.80	12.50
2016RAY03	812	10.08	1.71	0.060	3.56	14.48
2016WES04	1643	9.93	1.07	0.026	3.05	12.95
2016WEY05	835	11.68	1.43	0.049	3.80	15.70
2017ARB01	795	8.96	1.79	0.063	1.78	13.97
2017GFC02	560	4.89	1.37	0.058	1.02	9.14
2017IFC08	673	11.47	1.51	0.058	3.30	16.00
2017RAY05	796	9.25	1.53	0.054	3.05	13.21
2017RAY06	797	8.26	1.61	0.057	1.27	12.70
2017WES04	786	8.60	1.52	0.054	1.27	12.45
2017WES07	1596	9.52	1.16	0.029	4.57	13.21
2017WEY09	807	9.92	2.10	0.074	3.20	15.90

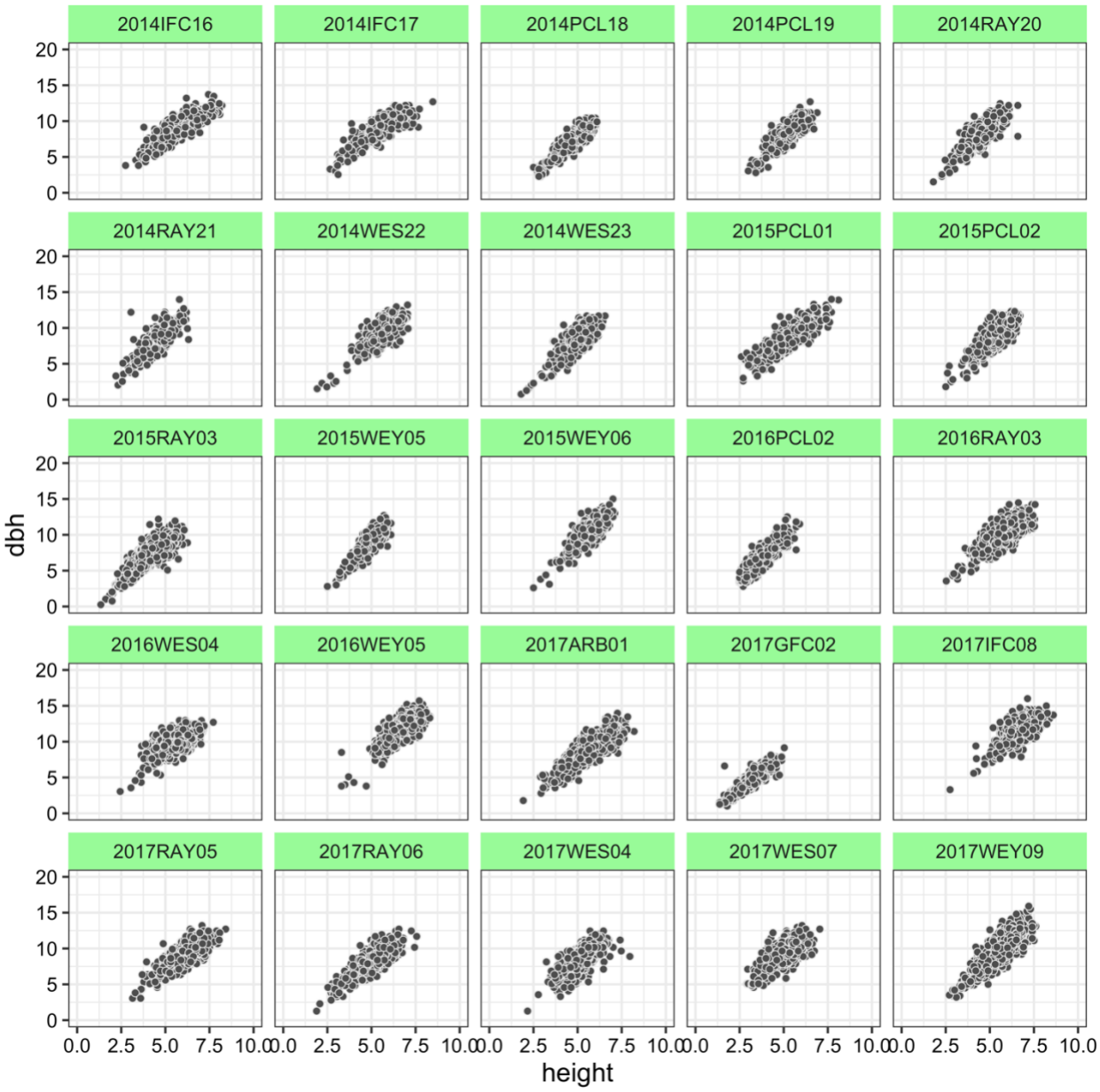

Figure A.2.1 Scatter plot showing the Height - DBH distribution at each site.

Appendix B. Codes Used in Chapter 2

Appendix B.1. ASReml Codes Used in Chapter 2

```
!WORKSPACE 32 !ARGS !RENAME 2 !OUTFOLDER
C:\Users\sinemsatiroglu\Dropbox\sinem\multisite\animal\out
C:\Users\nasir\Dropbox\sinem\multisite\animal\out

#!DOPATH $A

Title: Fourth-Cycle data analysis

 tree  !A
 female  !P
 male  !P
 family  !A !LL 17
 test  !A 25
 rep  *
 row  *
 col  *
 height
 dbh
 vol
 rust
 strt
 forkram
 HTm
 DBHcm

!FOLDER C:\Users\sinemsatiroglu\Dropbox\sinem\multisite\animal
# pedigree
pedigree_family.csv !SKIP 1 !ALPHA #!SORT #!DIAG !GIV 2
# data animal
cdat_fsonly.csv !SKIP 1 !DOPART $A !MVINCLUDE !MAXIT 100

#TABULATE DBHcm HTm ~ test !STATS !DEC 2
```

###ASReml Codes for Estimating Genetic Parameter in Height####

###ASReml Codes for Estimating Genetic Parameter with IID structure in Height

```
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
# height analysis
# IID treeID, IID treeID.site
# IID cross, IID cross.site
```

```
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
!PART 1
!MAXIT 100 !CONTINUE !MSV !FCON #!NODISPLAY #!AISING
$B ~ mu test,
     !r idh(test).rep, # rep variance for each site
        idh(test).rep.row, # row variance for each site
        idh(test).rep.col, # col variance for each site
        nrm(tree),
        nrm(tree).test,
        family,
        test.family
        residual sat(test).id(units) # residual variance for
each site
VPREDICT !DEFINE
```

###ASReml Codes for Estimating Genetic Parameter with CORUH structure in Height

```
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
# height analysis
# CORUH treeID nested within site
# IID cross, IID cross.site
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
!PART 2
!MAXIT 100 !CONTINUE !MSV !FCON #!NODISPLAY #!AISING
$B ~ mu test,
     !r idh(test).rep, # rep variance for each site
        idh(test).rep.row, # row variance for each site
        idh(test).rep.col, # col variance for each site
        coruh(test).nrm(tree),
        family,
        test.family
        residual sat(test).id(units) # residual variance for
each site
VPREDICT !DEFINE
```

###ASReml Codes for Estimating Genetic Parameter with XFA1 structure in Height

```
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
# height analysis - SELECTED FINAL MODEL
# XFA1 treeID nested within site
# IID cross, IID cross.site
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
!PART 3
!MAXIT 100 !CONTINUE !MSV !FCON #!NODISPLAY #!AISING
$B ~ mu test,
     !r idh(test).rep,     # rep variance for each site
```

```
        idh(test).rep.row, # row variance for each site
        idh(test).rep.col, # col variance for each site
        xfa1(test).nrm(tree),
        family,
        test.family
 residual sat(test).id(units) # residual variance for each site
VPREDICT !DEFINE
```

Heritability Calculation for Height using XFA1 structured model

```
# Convert XFA to US form
V Tree xfa1(test).nrm(tree);xfa1(test)
# Sum of all sites error variances
F  ErrSum
3+4+5+6+7+8+9+10+11+12+13+14+15+16+17+18+19+20+21+22+23+24+25+26
+27
# Mean error variance across all sites
F  ErrAve    ErrSum*0.04 # 1/25 = 0.04
# Error variance for family means
F  ErrMHS      ErrSum*0.0000011 # t = 25, nh = 5, p=294,  1/25^2
* 1/5 * 1/293 = 0.0000011
# Error variance for full sib family means
F  ErrMFS      ErrSum*0.00032 # t = 25, nh = 5,  1/25^2 * 1/5 =
0.00032
# Sum of all sites tree variances
F  TreeSum
153+155+158+162+167+173+180+188+197+207+218+230+243+257+272+288+
305+323+342+362+383+405+428+452+477
# Mean tree variance across all sites
F  TreeAve   TreeSum*0.04
# Additive variance, tree variance x average additive by site
correlation
F  AddVar    TreeAve*0.690233 # 0.69 is the average pairwise
additive correlation
# Numerator for HS family mean
F  HSnum     AddVar*0.25
# Family variance
F  Fam       1
# Family by site variance
F  FamGE     2
# Dominance genetic variance
F  DomVar    Fam*4.0
# Genetic variance
F  GenVar    AddVar+DomVar
# Dominance by environment denominator
F DomGE      DomVar+FamGE
# Phenotype variances
```

```
# Phenotypic variance for individual tree
F  Phen_i     TreeAve + Fam + FamGE + ErrAve
# Half-sib family mean phenotypic variance components
F  C1         TreeAve*0.04
F  C2         TreeAve*0.6626208
F  C3         Fam*0.00341297 # 1/293, p-1 = 294-1
F  C4         FamGE*0.00013652
# Half-sib family mean phenotypic variance
F  Phen_hs    485+493+494+495+480
# Full-sib family mean phenotypic variance components
F  J1         AddVar*0.5
F  J2         497*0.96
F  J3         Fam
F  J4         FamGE*0.04
# Full-sib family mean phenotypic variance
F  Phen_fs    497+498+499+500+481
# Individual narrowsense heritability
H  h_i        AddVar  Phen_i
# Individual broadsense heritability
H  H_i        GenVar  Phen_i
# Half-sib family mean heritability
H  H_hs       485  Phen_hs
# Full-sib family mean heritability
H  H_fs       497      Phen_fs
# Dominance by environment correlation
H rD          DomVar  DomGE
# Additive GxE correlations
R rA Tree
```

###ASReml Codes for Estimating Genetic Parameter with IID structure in DBH

```
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
# dbh analysis
# IID treeID, IID treeID.site
# IID cross, IID cross.site
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
!PART 5
!MAXIT 100 !CONTINUE !MSV !FCON #!NODISPLAY #!AISING
$B ~ mu test,
     !r idh(test).rep, # rep variance for each site
        idh(test).rep.row, # row variance for each site
        idh(test).rep.col, # col variance for each site
        nrm(tree),
        nrm(tree).test,
        family,
        test.family
```

```
 residual sat(test).id(units) # residual variance for each site
VPREDICT !DEFINE
```

###ASReml Codes for Estimating Genetic Parameter with CORUH structure in DBH

```
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
# dbh analysis
# CORUH treeID nested within site
# IID cross, IID cross.site
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
!PART 6
!MAXIT 100 !CONTINUE !MSV !FCON #!NODISPLAY #!AISING
$B ~ mu test,
     !r idh(test).rep, # rep variance for each site
        idh(test).rep.row, # row variance for each site
        idh(test).rep.col, # col variance for each site
        coruh(test).nrm(tree),
        family,
        test.family
  residual sat(test).id(units) # residual variance for each site
VPREDICT !DEFINE
```

###ASReml Codes for Estimating Genetic Parameter with XFA1 structure in DBH

```
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
# dbh analysis. This is used as the final model
# XFA1 treeID nested within site
# IID cross, IID cross.site
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
!PART 4
!MAXIT 100 !CONTINUE !MSV !FCON #!NODISPLAY #!AISING
$B ~ mu test,
     !r idh(test).rep, # rep variance for each site
        idh(test).rep.row, # row variance for each site
        idh(test).rep.col, # col variance for each site
        xfa1(test).nrm(tree),
        family,
        test.family
 residual sat(test).id(units) # residual variance for each site
VPREDICT !DEFINE
```

Heritability Calculation for DBH using XFA1 structured model

```
# Convert XFA to US form
V Tree xfa1(test).nrm(tree);xfa1(test)
```

```
# Sum of all sites error variances
F  ErrSum
3+4+5+6+7+8+9+10+11+12+13+14+15+16+17+18+19+20+21+22+23+24+25+26
+27
# Mean error variance across all sites
F  ErrAve    ErrSum*0.04 # 1/25 = 0.04
# Error variance for family means
F  ErrMHS      ErrSum*0.0000007 # t = 25, nh = 8.3, p=294,
1/25^2 * 1/8.3 * 1/293 = 0.0000007
# Error variance for full sib family means
F  ErrMFS      ErrSum*0.0002 # t = 25, nh = 8.3,  1/25^2 * 1/8.3
= 0.0002
# Sum of all sites tree variances
F  TreeSum
153+155+158+162+167+173+180+188+197+207+218+230+243+257+272+288+
305+323+342+362+383+405+428+452+477
# Mean tree variance across all sites
F  TreeAve   TreeSum*0.04
# Additive variance, tree variance x average additive by site
correlation
F  AddVar    TreeAve*0.56 # 0.56 is the average pairwise
additive correlation
# Numerator for HS family mean
F  HSnum     AddVar*0.25
# Family variance
F  Fam       1
# Family by site variance
F  FamGE     2
# Dominance genetic variance
F  DomVar    Fam*4.0
# Genetic variance
F  GenVar    AddVar+DomVar
# Dominance by environment denominator
F DomGE      DomVar+FamGE
# Phenotype variances
# Phenotypic variance for individual tree
F  Phen_i    TreeAve + Fam + FamGE + ErrAve
# Half-sib family mean phenotypic variance components
F  C1        TreeAve*0.04
F  C2        TreeAve*0.5376
F  C3        Fam*0.00341297 # 1/293, p-1 = 294-1
F  C4        FamGE*0.00013652
# Half-sib family mean phenotypic variance
F  Phen_hs   485+493+494+495+480
# Full-sib family mean phenotypic variance components
F  J1        AddVar*0.5
F  J2        497*0.96
```

```
F  J3        Fam
F  J4        FamGE*0.04
# Full-sib family mean phenotypic variance
F  Phen_fs   497+498+499+500+481
# Individual narrowsense heritability
H  h_i       AddVar  Phen_i
# Individual broadsense heritability
H  H_i       GenVar  Phen_i
# Half-sib family mean heritability
H  H_hs      485  Phen_hs
# Full-sib family mean heritability
H  H_fs      497      Phen_fs
# Dominance by environment correlation
H rD         DomVar  DomGE
# Additive GxE correlations
R rA Tree
```

Appendix B.2. R Codes Used in Chapter 2

R code for Figure 2.1

```
# clean all from the environment
rm(list = ls())
library(tidyverse)

# set working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

# read data
dat <- read.csv("/Users/ sinemsatiroglu/Dropbox/sinem/explanatory analysis/cdat_fullsib.csv",
header = TRUE)

# The path for map
gpath <- "/Users/ sinemsatiroglu/Dropbox/sinem/explanatory analysis/fullsib data"

###Creating map to show test location
##Installing packages
library(maps)
library(mapdata)
library(ggmap)
library(ggplot2)
library(cowplot)
library(tidyverse)
library(rgdal)
library(broom)
library(ggrepel)


# Read site locations data
locations <- read.csv("/Users/ sinemsatiroglu/Dropbox/sinem/explanatory analysis/fullsib
data/site map/map/cycle4_test_locations.csv")

# Data correction
locations$longitude <- ifelse(locations$longitude>0 , locations$longitude* -1 ,
locations$longitude)

# Extract test sites from cycle4 data
```

```
sites <- data.frame(table(dat$test))

names(sites)[names(sites)=="Var1"] <- "test_id"
names(sites)[names(sites)=="Freq"] <- "n"

# Merger locations and test sites
sites <- merge(locations, sites , by.y = "test_id" , all.x = FALSE)

### Layer of map data
ptaeda <- readOGR(dsn="/Users/nasir/Dropbox/sinem/explanatory analysis/fullsib data/site
map/map" , layer = "pinutaed")


ptaeda2=data.frame(ptaeda)
glimpse(ptaeda)

# Extract the data
ptaeda_tidy <- tidy(ptaeda)

# Extract states info
states <- map_data("state")
states$region <- as.factor(states$region)

# Extract southeast states only
southeast <- subset(states, region %in% c(c("arkansas" ,
"alabama","florida","georgia","maryland",
                                  "mississippi","north carolina","delaware",
                                  "south carolina","tennessee","virginia", "texas", "louisiana") ))

# ggplot

sitemap <- ggplot() + theme_classic() +
  geom_polygon(data=ptaeda, aes(x = long, y=lat, group=group),
               fill = "palegreen", color= "NA") +
  geom_polygon(data=southeast, aes(x = long, y = lat, group = group),
               fill = "NA", color = "black", size=.3) +
  geom_point(data=sites, aes(longitude , latitude), color="purple",
             size=2 , alpha= .6) +
  geom_jitter()+
  #geom_text_repel(aes(label = sites),
                  #box.padding   = 0.35,
                  #point.padding = 0.5,
                  #segment.color = 'grey50') +
  theme_void()

#geom_jitter(width = 0.1, height = 0.1)
```

sitemap

```
# save the map
# save figure 1
full.path = paste(gpath,"/sites-map.pdf", sep="")
pdf(full.path, width = 8, height = 5.5)
sitemap
dev.off()
```

R code for Figure 2.2

```
# set working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

# clean all from the environment
rm(list = ls())

# load libraries
library(tidyverse)
library(reshape2)
library(stringr)
library(plyr)


# read data and pedigree
dat <- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/explanatory analysis/cdat_fullsib.csv",
header = TRUE)
ped <- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/explanatory analysis/cparent.csv", header
= TRUE)
attach(dat)

# subset the data
data <- dat %>%
  select(female,male,test)

##This function determines the number of observations occurring across all combinations
of the factor levels of two factors
##df1 is an empty data frame, the row names being the factor levels of factor 1 (unique
parent names), the colnames being the factor levels of factor 2 (site)
##df2 is the a three column table: col1 = parentID, col2 = site, col3 = #observations
##for example, df2 column 1 would be “parent1” and df2 column 2 would be “test”; df2
column 3 may be # trees (must be >0)

data$year = dat$test
data$year <- as.character(data$year)
```

```
data$year <- str_sub(data$year,1,nchar(data$year)-5)
data <- as.data.frame(data)

# subset parent 1
female<-data[,-2]

# subset paretnt 2
male<-data[, -1]

names(female)[names(female)=="female"]<-"Parent"
names(male)[names(male)=="male"]<-"Parent"
parent <-as.data.frame(rbind(female,male))
unique(parent)

dfA <- matrix(nrow = length(unique(parent$Parent)),
              ncol = length(unique(parent$test)))
dfA <- as.data.frame(dfA)
row.names(dfA) <- unique(parent$Parent)
names(dfA) <- unique(parent$test)

dfB <- aggregate(year ~ Parent + test , FUN = length, parent)
names(dfB)<- c("parent","test","numobs")

DesignMatrix<-function(df1,df2){
  factor1<-as.vector(as.character(rownames(df1)))
  factor2<-as.vector(as.character(colnames(df1)))
  factorlevels<-as.vector(as.character(unique(paste(df2[,1],df2[,2]))))
  output<-data.frame(matrix(NA,nrow=nrow(df1),ncol=ncol(df1),dimnames=dimnames(df1)))
  for(i in 1:nrow(output)){
    for(j in 1:ncol(output)){
      test=paste(row.names(df1[i,]),colnames(df1[j]))
      if(test %in% factorlevels){
        output[i,j]=1} else{output[i,j]=0}
    }
  }
  return(output)
}

dm <- DesignMatrix(dfA, dfB)

cm<-t(as.matrix(dm)) %*% as.matrix(dm)

cm <- cm[rownames(cm)[order(rownames(cm))], colnames(cm)[order(colnames(cm))] ]

# cm is the connection matrix
colnames(cm)
```

```
# hack - replace the "." with "-" - not sure where it is coming from
colnames(cm) <- sub(".","",colnames(cm))
rownames(cm) <- sub(".","",rownames(cm))


# prepare to plot
# make a function to get upper triangle
get_upper_tri <- function(cormat){
  cormat[lower.tri(cormat, diag = F)]<- NA
  return(cormat)
}
upper_tri <- get_upper_tri(cm)

# melt the upper triangle

melted_cormat <- melt(upper_tri, na.rm = TRUE)

# import the series for panel header names
melted_cormat$Ser1 <- data$year[match(melted_cormat$Var1, data$test)]
melted_cormat$Ser2 <- data$year[match(melted_cormat$Var2, data$test)]

# import the test id for axis tick labels
melted_cormat$Test1 <- data$test[match(melted_cormat$Var1, data$test)]
melted_cormat$Test2 <- data$test[match(melted_cormat$Var2, data$test)]

# order factor levels to arrange plot correctly (tricky)
melted_cormat$Ser2 <- factor(melted_cormat$Ser2, levels = c("2014","2015","2016", "2017"))
melted_cormat$Ser1 <- factor(melted_cormat$Ser1, levels = c("2017","2016","2015", "2014"))

# make the plot
# open pdf file

tiff('genetic connection3.tiff', units="in", width=6, height=6, res=500)
pdf(file = "/Users/sinemsatiroglu/Dropbox/sinem/explanatory analysis/genetic_connection3.pdf",
height = 6, width = 6)
ggplot(data = melted_cormat, aes(Test2, Test1, fill = value))+
  geom_tile(color = "white", size=.3)+
  facet_grid(Ser1 ~ Ser2, drop = T, space="free", scales = "free") +
  scale_fill_gradientn(colours =  c("khaki","green","green2","green4"),  # adding more "green"
pushes red values down
                       #breaks = c(0, 15, 100, 300), #limits=c(0,300),  ## edit these to customize color
range
                       #labels= c("0", "15", "100", "300"),
                       guide = guide_colourbar(nbin = 100),
                       na.value = NA, name="# of shared parents") +
```

```
  geom_text(aes(Test2, Test1, label = value), color = "black", size = 2) +
  theme(plot.title = element_text()) +
  theme(axis.text = element_text(color='black', size=9),
        axis.ticks = element_line(size=.4),
        axis.text.x = element_text(angle = 90, vjust=.5)) +
  theme(axis.title = element_blank(),
        panel.grid = element_blank(),
        panel.background = element_blank())+
  theme(legend.position=c(.15,.75),
        legend.key.width=unit(1,"cm"),
        legend.key.height=unit(1,"cm"),
        panel.border = element_rect(color = NA, fill = NA, size = .3),
        panel.spacing = unit(.3, "mm"),
        strip.background = element_rect(fill="cyan"))

dev.off()
```

R code for Figure 2.3, 2.4 and Supplementary Figure 2.1

```
# set working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

# clean all from the environment
rm(list = ls())

# load libraries
library(tidyverse)
library(reshape2)
library(stringr)
library(plyr)

# read data and pedigree
dat <- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/explanatory analysis/cdat_fullsib.csv",
header = TRUE)
ped <- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/explanatory analysis/cparent.csv", header
= TRUE)
attach(dat)

## test site statistics for height
ht <- ddply(dat, c("test", "test_id"), summarise,
            N = length(female),
            Height = round(mean(height, na.rm = TRUE), digits = 2),
            HeightSD = round(sd(height, na.rm = TRUE), digits = 2),
            HeightSE = round(HeightSD/sqrt(N), digits = 3),
            HeightMin = round(min(height, na.rm = TRUE), digits = 2),
```

```
        HeightMax = round(max(height, na.rm = TRUE), digits = 2))

write.table(ht, file = "heightinfo.csv", sep = ",", col.names = NA,
        qmethod = "double")


## test site statistics for dbh
dbh <- ddply(dat, c("test", "test_id"), summarise,
          N = length(female),
          DBH = round(mean(dbh, na.rm = TRUE), digits = 2),
          DBHSD = round(sd(dbh, na.rm = TRUE), digits = 2),
          DBHSE = round(DBHSD/sqrt(N), digits = 3),
          DBHMin = round(min(dbh, na.rm = TRUE), digits = 2),
          DBHMax = round(max(dbh, na.rm = TRUE), digits = 2))

write.table(dbh, file = "dbhinfo.csv", sep = ",", col.names = NA,
        qmethod = "double")


# boxplot of height for each site
pdf(file = "/Users/sinemsatiroglu/Dropbox/sinem/explanatory analysis/height.pdf", height = 6,
width = 5)
tiff("height.tiff", height = 6, width = 5, res = 500, units = 'in')
ggplot(dat, aes(x=test, y=height))+ theme_bw()+
  geom_boxplot(color='blue', outlier.color = 'cornsilk3', outlier.size = 1)+
  stat_summary(fun=mean, geom="point", shape=20, size=3, color="red", fill="red4") +
  coord_flip()+
  scale_y_continuous(limits = c(0,10))+
  labs(x="Test site", y="Height (m)")+
  theme(axis.text = element_text(color='black', size=9),
        axis.title = element_text(size = 12))
dev.off()

# boxplot of dbh for each site
pdf(file = "/Users/sinemsatiroglu/Dropbox/sinem/explanatory analysis/dbh.pdf", height = 6,
width = 5)
tiff("dbh.tiff", height = 6, width = 5, res = 500, units = 'in')
ggplot(dat, aes(x=test, y=dbh))+ theme_bw()+
  geom_boxplot(color='blue', outlier.color = 'cornsilk3', outlier.size = 1)+
  stat_summary(fun=mean, geom="point", shape=20, size=3, color="red", fill="red4") +
  coord_flip()+
  scale_y_continuous(limits = c(0,20))+
  labs(x="Test site", y="DBH (in)")+
  theme(axis.text = element_text(color='black', size=9),
        axis.title = element_text(size = 12))
dev.off()
```

```
# eliminate the extremes

dat <- mutate(dat, height=ifelse(height<1, NA, height),
            dbh=ifelse(c(dbh<5 & height >6), NA, dbh),
            dbh=ifelse(c(dbh < 6.7 & height< 1.65), NA, dbh))

# height-dbh scatterplots
pdf(file = "/Users/sinemsatiroglu/Dropbox/sinem/explanatory analysis/height_dbh.pdf", height =
6, width = 5)
tiff("ht-dbh.tiff", height = 7, width = 7, res = 500, units = 'in')
ggplot(dat, aes(x=height, y=dbh))+ theme_bw()+
  geom_point(pch=21, color='gray90', size=1.5, fill='gray30', stroke=.3)+
  facet_wrap(~test)+
  scale_y_continuous(limits = c(0,20))+
  scale_x_continuous(limits = c(0,10))+
  theme(axis.text = element_text(color='black', size=9),
        axis.title = element_text(size = 12),
        strip.background = element_rect(fill='palegreen',color=NA))
dev.off()

### R code for Figure 2.5

### Methods: In order to create the heatmap, first dendrogram was
### created using agglomerative heirarchical clustering "hclust"
### function in R. Then heatmap with the dendrogram was created
### using "heatmap.2" function of gplots package as follows.

#Remove everything from memory
rm(list=ls())

#Setting working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

# The path for map
#gpath <- "/Users/sinemsatiroglu/Dropbox/sinem/figures/chapter 1"

library(gplots)
library(tidyverse)

data <- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/explanatory analysis/fullsib data/rA
matrix_height.csv",
              header = TRUE, row.names = 1)

# remove X from colnames
names(data)<-gsub("\\X","",names(data))
```

```
## convert the data frame into numeric matrix

mat<- data.matrix(data, rownames.force = NA)
mat
cd = dist(mat)
cd
hc = hclust(cd, method = "complete")
hc
cdt = dist(t(mat))
cdt
hcc = hclust(cdt, method = "complete")
hcc

pdf(file = "/Users/sinemsatiroglu/Dropbox/sinem/figures/chapter 1/correlation_height.pdf",
height = 15, width = 15)
tiff('correlation_height.tiff', units="in", width=11, height=10.5, res=500)
heatmap.2(mat, Rowv = as.dendrogram(hc), Colv = as.dendrogram(hcc), dendrogram = "row",
        density.info = 'histogram', trace = 'none', cexRow = 1, cexCol = 1,
        rowsep = c(1:25), colsep = c(1:25),key.xlab = 'GxE correlation',
        sepcolor = "NA", sepwidth = c(.001,.001), key.title = "",
        key = TRUE, key.par = list(cex=1.3), keysize = 1.7,
        col = colorRampPalette(c('yellow','lightblue','blue')),
        notecex = 1.1, notecol = 'white',margins=c(6,6))
dev.off()
```

R code for Figure 2.6

```
### Methods: In order to create the heatmap, first dendrogram was
### created using agglomerative heirarchical clustering "hclust"
### function in R. Then heatmap with the dendrogram was created
### using "heatmap.2" function of gplots package as follows.

#Remove everything from memory
rm(list=ls())

#Setting working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

library(gplots)

data <- read.csv("/Users/ sinemsatiroglu/Dropbox/sinem/explanatory analysis/fullsib data/rA
matrix_dbh.csv", header = TRUE, row.names = 1)

# remove X from colnames
names(data)<-gsub("\\X","",names(data))
```

convert the data frame into numeric matrix

```
mat<- data.matrix(data, rownames.force = NA)
mat
cd = dist(mat)
cd
hc = hclust(cd, method = "complete")
hc
cdt = dist(t(mat))
cdt
hcc = hclust(cdt, method = "complete")
hcc

pdf(file = "/Users/sinemsatiroglu/Dropbox/sinem/figures/chapter 1/correlation_dbh.pdf", height
= 15, width = 15)
tiff('correlation_dbh.tiff', units="in", width=11, height=10.5, res=500)
heatmap.2(mat, Rowv = as.dendrogram(hc), Colv = as.dendrogram(hcc), dendrogram = "row",
        density.info = 'histogram', trace = 'none', cexRow = 1, cexCol = 1,
        rowsep = c(1:25), colsep = c(1:25),key.xlab = 'GxE correlation',
        sepcolor = "NA", sepwidth = c(.001,.001), key.title = "",
        key = TRUE, key.par = list(cex=1.3), keysize = 1.7,
        col = colorRampPalette(c('yellow','lightblue','blue'))(n = 50),
        notecex = 1.1, notecol = 'white',margins=c(6,6))
dev.off()
```

R code for Figure 2.7

```
#Remove everything from memory
rm(list=ls())

#Setting working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

library(tidyverse)
library(reshape)
library(reshape2)
library(GGally)
library(extrafont)
loadfonts(device = "win")

dat <- read.csv(file = 'siteinfo and loadings.csv', header = TRUE)
# read loading data sorted by trait
load <- read.csv(file = 'loadings.csv', header = TRUE)

# read in climate data
```

```
clime <- read.csv(file = 'sitedifferences.csv', header = TRUE)
# read in climate data organized for scatterplots
clmdat <- read.csv(file = 'GEbyEnvr.csv', header = TRUE)


# calculate correlation between loadings and environmental variables
corrs2 <- clmdat %>%
  group_by(Trait1,Trait2) %>%
  summarise(r = round(cor(AddGE,Value),digits = 2))

# create a new column
corrs2$r2 <- paste('r =',corrs2$r)
corrs2$xpos <- c(6,12,6,12)

#plot
tiff('GE by Env correlations.tiff', units="in", width=7, height=5, res=500)
ggplot(clmdat, aes(x=Value, y=AddGE))+ theme_bw()+
  geom_point(color='dodgerblue')+
  geom_smooth(method = 'lm', formula = y~x, color='black', size=.8)+
  geom_text(mapping = aes(x=xpos, y=-.4, label =r2),data = corrs2,
            color='red2', fontface='italic', size=3.5)+
  facet_grid(Trait1~Trait2, scales = 'free_x')+
  labs(y='Additive GxE correlation')+
  theme(axis.text = element_text(color='black', size=10),
        axis.title.x = element_blank(),
        axis.title = element_text(size=12),
        panel.border = element_rect(size=.4),
        strip.text = element_text(size=12))
dev.off()

# Fitting linear models
# Height
mod.ht <- lm(rA_HT ~ LatLongDiffSum+MWTDiff+LatLongDiffSum:MWTDiff, data = clime)
summary(mod.ht)
anova(mod.ht)

# DBH
mod.dbh <- lm(rA_DBH ~ LatLongDiffSum+MWTDiff+LatLongDiffSum:MWTDiff, data =
clime)
summary(mod.dbh)
anova(mod.dbh)
```

Appendix C. Codes Used in Chapter 3

Appendix C.1. ASReml Codes Used in Chapter 3

```
!WORKSPACE 32 !ARGS !RENAME 2 !OUTFOLDER
C:\Users\sinemsatiroglu\Dropbox\sinem\multisite\animal\out
C:\Users\nasir\Dropbox\sinem\multisite\animal\out

#!DOPATH $A

Title: Fourth-Cycle data analysis

 tree  !A
 female  !P
 male  !P
 family  !A !LL 17
 test  !A 25
 rep  *
 row  *
 col  *
 height
 dbh
 vol
 rust
 strt
 forkram
 HTm
 DBHcm

!FOLDER C:\Users\sinemsatiroglu\Dropbox\sinem\multisite\animal
# pedigree
pedigree_family.csv !SKIP 1 !ALPHA #!SORT #!DIAG !GIV 2
# data animal
cdat_fsonly.csv !SKIP 1 !DOPART $A !MVINCLUDE !MAXIT 100
```

###ASReml Codes for Estimating Genetic Parameter in Fusiform Rust Disease Incidence

```
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
# rust analysis
# IID treeID, IID treeID.site
# IID cross, IID cross.site
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
!PART 7
!CONTINUE !MSV !FCON !NODISPLAY #!AISING
$B !BIN !LOGIT ~ mu test,
     !r test.rep,
        rep.row,
        rep.col,
        female and(male) ,
        female.test -male.test and(male.test),
        family,
        test.family

predict female !average test
predict family !average test
VPREDICT !DEFINE
```

Heritability Calculation for Fusiform Rust Disease Incidence

```
# Error variances
F  Err       Residual*3.29
# Female variance
F  Fem       4
# Female by site variance
F  FemGE     5
# Family variance
F  Fam       6
# Family by site variance
F  FamGE     7
# Additive genetic variance
F  AddVar    Fem*4.0
# Dominance genetic variance
F  DomVar    Fam*4.0
# Genetic variance
F  GenVar    AddVar+DomVar
# Phenotypic variance for individual tree
F  Phen_i    Fem*2.0 + Fam + FemGE*2.0 + FamGE + Err
# Additive GxE correlation denominator
F  ADenom    AddVar+FemGE
# Dominance GxE correlation denominator
F  DDenom    DomVar+FamGE
# Genetic GxE correlation denominator
F  GDenom    AddVar+FemGE+DomVar+FamGE
# Half-sib family mean phenotypic variance components
```

```
F  C1         Fem # gcs var
F  C2         Fam*0.003413 # sca var/p-1, p = number of parents
F  C3         FemGE*0.04 # gcs by site/t, t = number of sites
F  C4         FamGE*0.000136519 # sca by site/t(p-1)
F  C5         Err*0.00001625 # error/tn(p-1) harmonic mean number
of trees per site = 8.4
# Half-sib family mean phenotypic variance
F  HSphen     C1+C2+C3+C4+C5
# Full-sib family mean phenotypic variance components
F  K1          Fem*2.0
F  K2          Fam
F  K3          FemGE*0.08 # 2*GCA by site/25
F  K4          FamGE*0.04 # SCA by site/25
F  K5          Err*0.00702 # Err/tn, n = 5.7
# Full-sib family mean phenotypic variance
F  FSphen      K1+K2+K3+K4+K5
# Individual narrowsense heritability
H  h_i        AddVar  Phen_i
# Individual broadsense heritability
H  H_i        GenVar  Phen_i
# Half-sib familymean heritability
H  H_hs       AddVar    HSphen
# Full-sib familymean heritability
H  H_fs       K1         FSphen
# Additive GxE correlation
H  rA         AddVar  ADenom
# Dominance GxE correlation
H  rD         DomVar  DDenom
# Genetic GxE correlation
H  rG         GenVar GDenom
# Ration between additive and dominance
H  ADratio    AddVar  DomVar
```

###ASReml Codes for Estimating Genetic Parameter in Stem Forking Incidence

```
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
# Forking analysis
### ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ###
!PART 8
!CONTINUE !MSV !FCON !NODISPLAY #!AISING
$B !BIN !LOGIT ~ mu test,
     !r test.rep,
        rep.row,
        rep.col,
        female and(male) ,
        female.test -male.test and(male.test),
        family,
```

```
        test.family

predict female !average test
predict family !average test
VPREDICT !DEFINE
```

Heritability Calculation for Stem Forking Incidence

```
# Error variances
F  Err       Residual*3.29
# Female variance
F  Fem       4
# Female by site variance
F  FemGE     5
# Family variance
F  Fam       6
# Family by site variance
F  FamGE     7
# Additive genetic variance
F  AddVar    Fem*4.0
# Dominance genetic variance
F  DomVar    Fam*4.0
# Genetic variance
F  GenVar    AddVar+DomVar
# Phenotypic variance for individual tree
F  Phen_i    Fem*2.0 + Fam + FemGE*2.0 + FamGE + Err
# Additive GxE correlation denominator
F  ADenom    AddVar+FemGE
# Dominance GxE correlation denominator
F  DDenom    DomVar+FamGE
# Genetic GxE correlation denominator
F  GDenom    AddVar+FemGE+DomVar+FamGE
# Individual narrowsense heritability
H  h_i       AddVar  Phen_i
# Individual broadsense heritability
H  H_i       GenVar  Phen_i
# Additive GxE correlation
H  rA        AddVar  ADenom
# Dominance GxE correlation
H  rD        DomVar  DDenom
# Genetic GxE correlation
H  rG        GenVar GDenom
```

```
# Ration between additive and dominance
H  ADratio   AddVar  DomVar
```

Appendix C.2. R Codes Used in Chapter 3

 # R codes for Figure 3.1

 # Create a boxplot showing rust incidence by family and population

#Remove everything from memory
rm(list=ls())

#Setting working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

library(tidyverse)
library(patchwork)

dat <- read.csv(file = 'cdat_fsonly.csv', header = TRUE)

fm <- dat %>%
  group_by(family) %>%
  summarise(rust = mean(rust, na.rm = TRUE),
            fork = mean(forkram, na.rm = TRUE))

 # convert to long format
fmlong <- gather(fm, trait, mean, rust:fork, factor_key=TRUE)

#plot
tiff('rustfork.tiff', units="in", width=6, height=7, res=500)
p1 <- ggplot(fm, aes(x=reorder(family,rust), y=rust)) + theme_bw()+
  geom_point(size=.8, color="black", shape=21, stroke=.1, fill='black') +
  geom_segment(aes(x=family, xend=family, y=0, yend=rust), color='blue', alpha=.7, size=.3)+
  labs(x = "", y = "Rust incidence (Prob)") +
  scale_y_continuous(expand = c(0,0), limits = c(0,1))+
  theme(axis.text=element_text(size=12, colour = "black"),
        axis.title=element_text(size=16,face="plain"),
        axis.title.x = element_text(vjust = -.4),
        axis.title.y = element_text(vjust = 1.5),

```
      axis.text.x = element_blank(),
      axis.ticks.x = element_blank(),
      panel.grid = element_blank())

p2 <- ggplot(fm, aes(x=reorder(family,fork), y=fork)) + theme_bw()+
  geom_point(size=.8, color="black", shape=21, stroke=.1, fill='black') +
  geom_segment(aes(x=family, xend=family, y=0, yend=fork), color='blue', alpha=.7, size=.3)+
  labs(x = "Family", y = "Forking incidence (Prob)") +
  scale_y_continuous(expand = c(0,0), limits = c(0,1))+
  theme(axis.text=element_text(size=12, colour = "black"),
      axis.title=element_text(size=16,face="plain"),
      axis.title.x = element_text(vjust = -.4),
      axis.title.y = element_text(vjust = 1.5),
      axis.text.x = element_blank(),
      axis.ticks.x = element_blank(),
      panel.grid = element_blank())
p1 + p2 + plot_layout(ncol=1)
dev.off()
```

 # R codes for Figure 3.2

```
#Remove everything from memory
#rm(list=ls())
library(tidyverse)
library(ggplot2)

#Setting working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

#getting data for parents
# this data get from solution file of GCA model 7

r1<- read.csv("/Users/nasir/Dropbox/sinem/explanatory analysis/rust/rust_female.csv", header =
TRUE)

colnames(r1)[1] <- 'Parent'

#calculating BV
# m= -0.08881

r1$BVlogit <- (r1$BLUP*2) + (-0.08881)

#converting BLUPs and BV to probabilities

#female
r1$BVprob <- exp(r1$BVlogit)/(1+exp(r1$BVlogit))
```

```
# bring in female and male IDs
f.m <- read.csv("/Users/nasir/Dropbox/sinem/explanatory analysis/rust/females and males.csv",
header = TRUE)

# we only need 294 females and males not grand parents
r1a <- left_join(f.m,r1)

rr <- subset(r1a,BVprob < 0.5)

#plot
tiff("rust_female_BV_prob2.tiff", height = 5, width = 7, res = 500, units = 'in')
ggplot(r1a, aes(x=reorder(Parent,BVprob), y=BVprob))+ theme_bw()+
  geom_linerange(data=r1, mapping=aes(x=reorder(Parent,BVprob), ymin=0, ymax=BVprob),
                 size=.3, color='blue', alpha=1)+
  labs(x='Parents', y='Breeding value (Prob)')+
  scale_y_continuous(expand = c(0,0), limits = c(0,1))+
  theme(
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.text = element_text(color='black',size=10),
        axis.title = element_text(size=12))+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black"))
dev.off()
```

```
# R codes for Figure 3.3

#Setting working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

library(tidyverse)

# female BLUPs
#fem <- read.csv(file = 'rust_female.csv', header = TRUE)
fem <- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/figures/chapter
3/rustGV/rust_female.csv", header = TRUE)
colnames(fem) <- c("Parent","femBLUP","femSE")
fem$Female <- fem$Parent
fem$Male <- fem$Parent
fem$mBLUP <- fem$femBLUP

# Family BLUPs
fam <- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/figures/chapter
3/rustGV/family_rust.csv", header = TRUE)
colnames(fam) <- c("Cross","famBLUP","famSE")
```

```
# copy cross column
fam$Cross2 <- fam$Cross

# create female and male columns
fam2 <- fam %>% separate(Cross2, c("Female","Male"), sep="_X_")

# merge female blups
femfam <- left_join(fam2,fem[,c(2,4)])
# merge male blups
femfam2 <- left_join(femfam,fem[,c(5,6)])
# calculate genetic value
femfam2$GV <- femfam2$famBLUP+femfam2$femBLUP+femfam2$mBLUP+(-0.08881)
# convert to prob
femfam2$GVprob <- exp(femfam2$GV)/(1+exp(femfam2$GV))

#plot
tiff("rust_family_GV_prob.tiff", height = 5, width = 7, res = 500, units = 'in')
ggplot(femfam2, aes(x=reorder(Cross,GVprob), y=GVprob))+ theme_bw()+
  geom_linerange(data=femfam2, mapping=aes(x=reorder(Cross,GVprob), ymin=0,
ymax=GVprob),
                 size=.3, color='blue', alpha=1)+
  labs(x='Family', y='Genetic value (Prob)')+
  scale_y_continuous(expand = c(0,0), limits = c(0,1))+
  theme(
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank(),
    axis.text = element_text(color='black',size=10),
    axis.title = element_text(size=12))+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black"))
dev.off()
```

R codes for Figure 3.4

```
#Setting working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

#Remove everything from memory
rm(list=ls())
library(tidyverse)
library(ggplot2)

#getting data for female
# this data get from solution file of GCA model 8
```

```
f1<- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/explanatory
analysis/forkram/forkram_female.csv", header = TRUE)
colnames(f1)[1] <- 'Parent'

#calculating BV
# m=-1.603

f1$BVlogit <- (f1$BLUP*2) + (-1.603)
f1$probBV <- exp(f1$BVlogit)/(1+exp(f1$BVlogit))

# bring in female and male IDs
f.m <- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/explanatory analysis/forkram/females
and males.csv", header = TRUE)
# we only need 294 females and males not grand parents
f1a <- left_join(f.m,f1)

#Plot
tiff("forkram_female_BV_prob2.tiff", height = 5, width = 7, res = 500, units = 'in')
ggplot(f1a, aes(x=reorder(Parent,probBV), y=probBV))+ theme_bw()+
  geom_linerange(data=f1, mapping=aes(x=reorder(Parent,probBV), ymin=0, ymax=probBV),
                 size=.3, color='blue', alpha=1)+
  labs(x='Parents', y='Breeding value (Prob)')+
  scale_y_continuous(expand = c(0,0), limits = c(0,0.5))+
  theme(
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank(),
    axis.text = element_text(color='black',size=10),
    axis.title = element_text(size=12))+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black"))
dev.off()
```

R codes for Figure 3.5

```
#Remove everything from memory
rm(list=ls())
library(tidyverse)
library(ggplot2)

#Setting working directory
setwd("/Users/sinemsatiroglu/Dropbox/sinem")

# female BLUPs
fem <- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/explanatory
analysis/forkram/forkram_female.csv", header = TRUE)
colnames(fem) <- c("Parent","femBLUP","femSE")
```

```
fem$Female <- fem$Parent
fem$Male <- fem$Parent
fem$mBLUP <- fem$femBLUP

# Family BLUPs
fam <- read.csv("/Users/sinemsatiroglu/Dropbox/sinem/explanatory
analysis/forkram/forkram_family.csv", header = TRUE)
colnames(fam) <- c("Cross","famBLUP","famSE")
# copy cross column
fam$Cross2 <- fam$Cross
# create female and male columns
fam2 <- fam %>% separate(Cross2, c("Female","Male"), sep="_X_")

# merge female blups
femfam <- left_join(fam2,fem[,c(2,4)])
# merge male blups
femfam2 <- left_join(femfam,fem[,c(5,6)])

# calculate genetic value
femfam2$GV <- femfam2$famBLUP+femfam2$femBLUP+femfam2$mBLUP+(-1.603)
# convert to prob
femfam2$GVprob <- exp(femfam2$GV)/(1+exp(femfam2$GV))

#plot
tiff("forkram_family_GV_prob.tiff", height = 5, width = 7, res = 500, units = 'in')
ggplot(femfam2, aes(x=reorder(Cross,GVprob), y=GVprob))+ theme_bw()+
  geom_linerange(data=femfam2, mapping=aes(x=reorder(Cross,GVprob), ymin=0,
ymax=GVprob),
                 size=.3, color='blue', alpha=1)+
  labs(x='Family', y='Genetic value (Prob)')+
  scale_y_continuous(expand = c(0,0), limits = c(0,0.5))+
  theme(
    axis.text.x=element_blank(),
    axis.ticks.x=element_blank(),
    axis.text = element_text(color='black',size=10),
    axis.title = element_text(size=12))+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black"))
dev.off()
```